IEEE NPSS Short Course

Radiation Detection and Measurement

November 10 – 11, 2002

2002 Nuclear Science Symposium
Norfolk, Virginia

Pulse Processing and Analysis

Helmuth Spieler

Physics Division
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

These course notes and additional tutorials at
http://www-physics.lbl.gov/~spieler

Table of Contents

1. Overview

Purpose of pulse processing and analysis systems:

1. acquire electrical signal from detector

   typically a short current pulse

2. tailor the time response (i.e. "shape" the output pulse) of the system to optimize

   - minimum detectable signal
   - energy measurement (magnitude of signal)
   - event rate
   - time of arrival (timing measurement)
   - insensitivity to detector pulse shape
   - some combination of the above

   Generally, these cannot be optimized simultaneously

   $\Rightarrow$ compromises

3. digitize the signal and store for subsequent analysis

Additional requirements, depending on specific application, e.g.

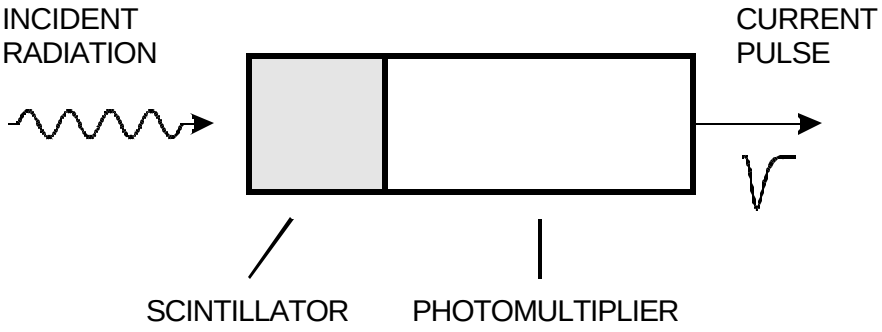
radiation resistance

low power
   portable systems
   large detector arrays, e.g. in HEP

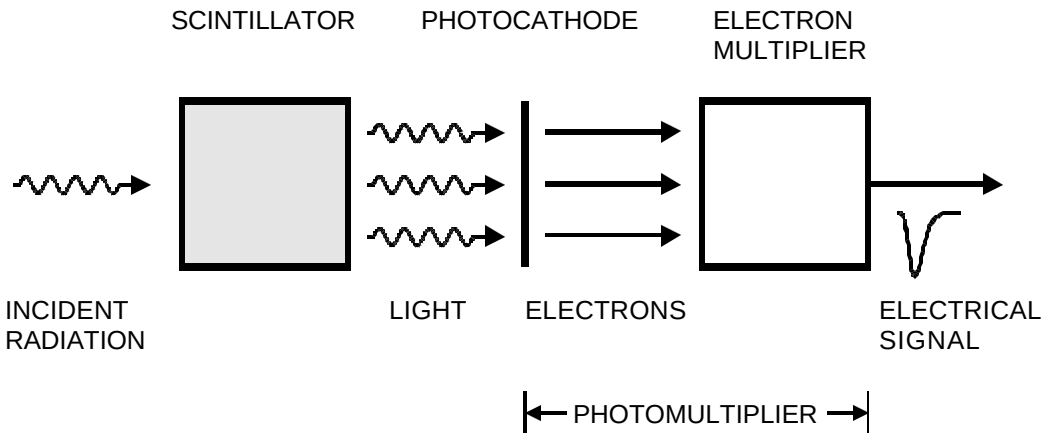
robustness

cost

A Typical Detector System – Scintillation Detector



Processes in Scintillator – Photomultiplier

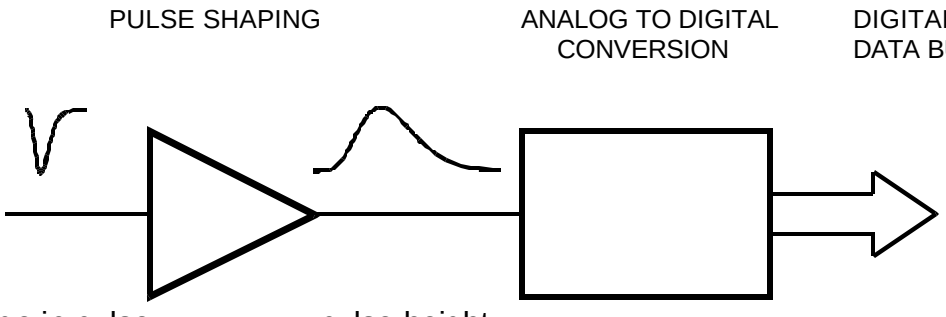


number of photons
∝ absorbed energy

number of photoelectrons
∝ absorbed energy

charge in pulse
∝ abs. energy

Signal Processing



charge in pulse
∝ abs. energy

pulse height
∝ absorbed energy

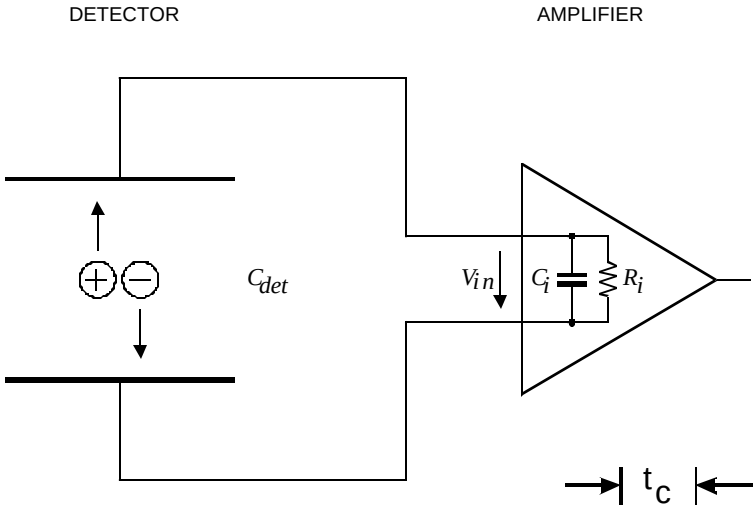
Ionization Chamber

All ionization chambers utilize the same principle:

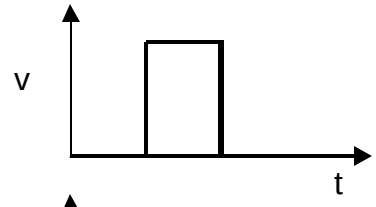
1. Particles deposit energy in an absorber and create mobile charge carriers (positive and negative charge pairs).

   in solids, liquids: electrons and holes
   in gases: electrons and ions

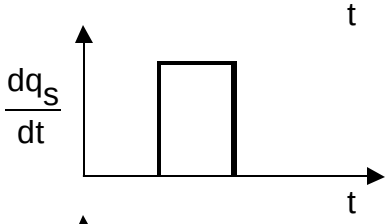
2. Electric field applied to detector volume sweeps charge carriers towards electrodes and induces a signal current



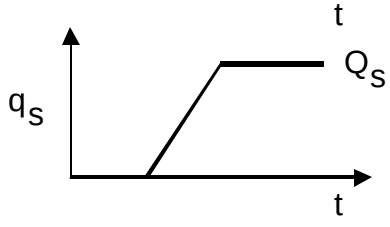
velocity of charge carriers



rate of induced charge on detector electrodes



signal charge



if R_i x (C_{det} + C_i) >> collection time t_c:

peak voltage at amplifier input $V_s = \frac{Q_s}{C_{\det} + C_i}$

The Signal

Any form of elementary excitation can be used to detect the radiation signal.

$$\text{Magnitude of signal} = \frac{\text{absorbed energy}}{\text{excitation energy}}$$

An electrical signal can be formed directly by ionization.

> Incident radiation quanta impart sufficient energy to individual atomic electrons to form electron-ion pairs (in gases) or electron-hole pairs (in semiconductors and metals).

Other detection mechanisms are

> Excitation of optical states (scintillators) $\rightarrow$ light intensity
>
> Excitation of lattice vibrations (phonons) $\rightarrow$ temperature
>
> Breakup of Cooper pairs in superconductors
>
> Formation of superheated droplets in superfluid He

Typical excitation energies

Ionization in gases	~30 eV
Ionization in semiconductors	1 – 10 eV
Scintillation	20 - 500 eV
Phonons	meV
Breakup of Cooper Pairs	meV

Precision of signal magnitude is limited by fluctuations

Two types of fluctuations

1. Fluctuations in signal charge for a given energy absorption in detector

   signal formed by many elementary excitations

$$\text{number of signal quanta} = \frac{\text{absorbed energy}}{\text{excitation energy}}$$

$$N = \frac{E}{E_i}$$

   Number of signal quanta fluctuates statistically.

$$\Delta N = \sqrt{FN}$$

   where F is the Fano factor (0.1 in Si, for example), so the energy resolution

$$\Delta E = E_i \Delta N = \sqrt{FEE_i} \quad \text{r.m.s.}$$

$$\Delta E_{FWHM} = 2.35 \times \Delta E_{rms}$$

2. Baseline fluctuations in the electronics

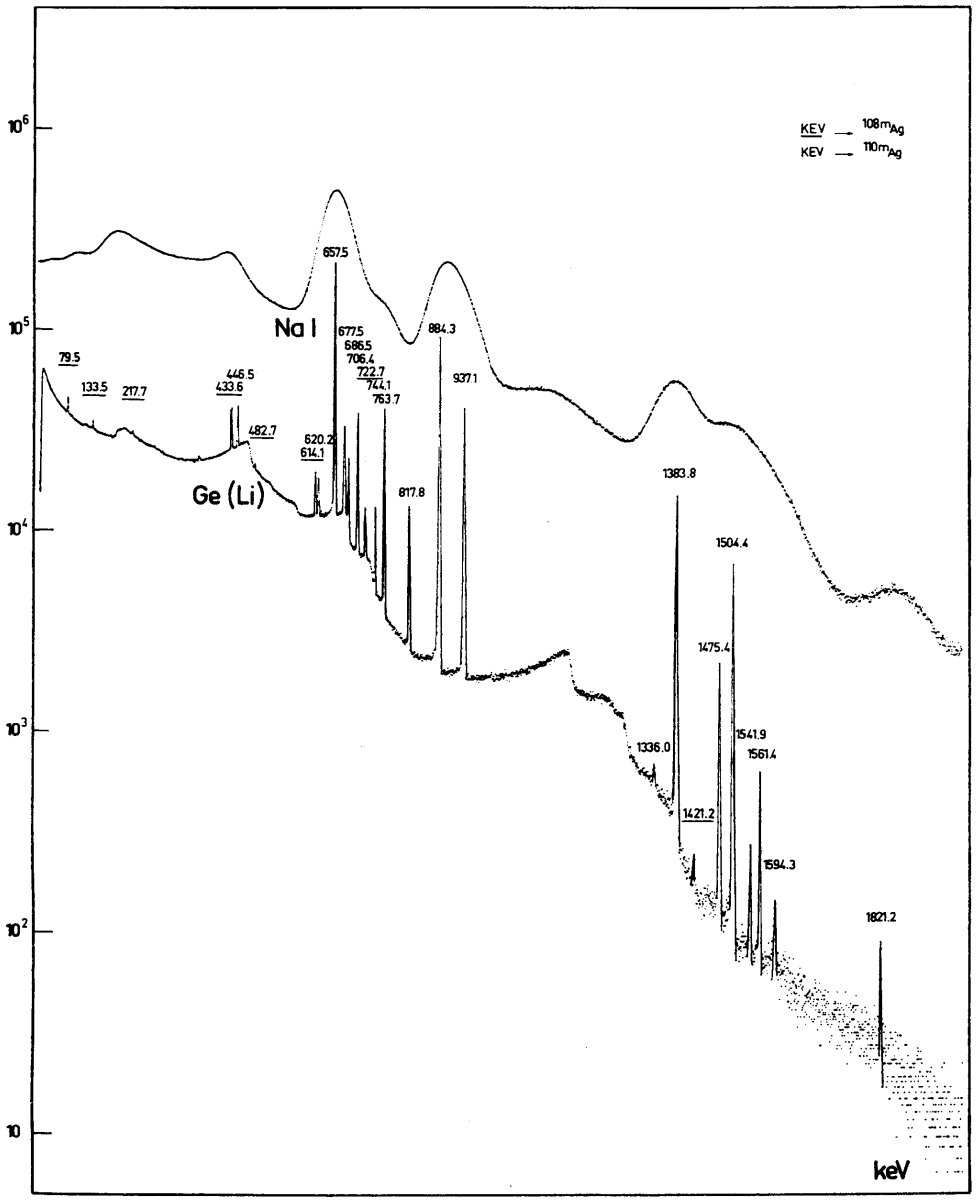
   "electronic noise"

The overall resolution is often the result of several contributions. Individual resolutions add in quadrature, for example

$$\Delta E = \sqrt{\Delta E_{fluc}^2 + \Delta E_{elec}^2}$$

If one contribution is 20% of the other, the overall resolution is increased by 10%.

Resolution of NaI(Tl) and Ge detectors



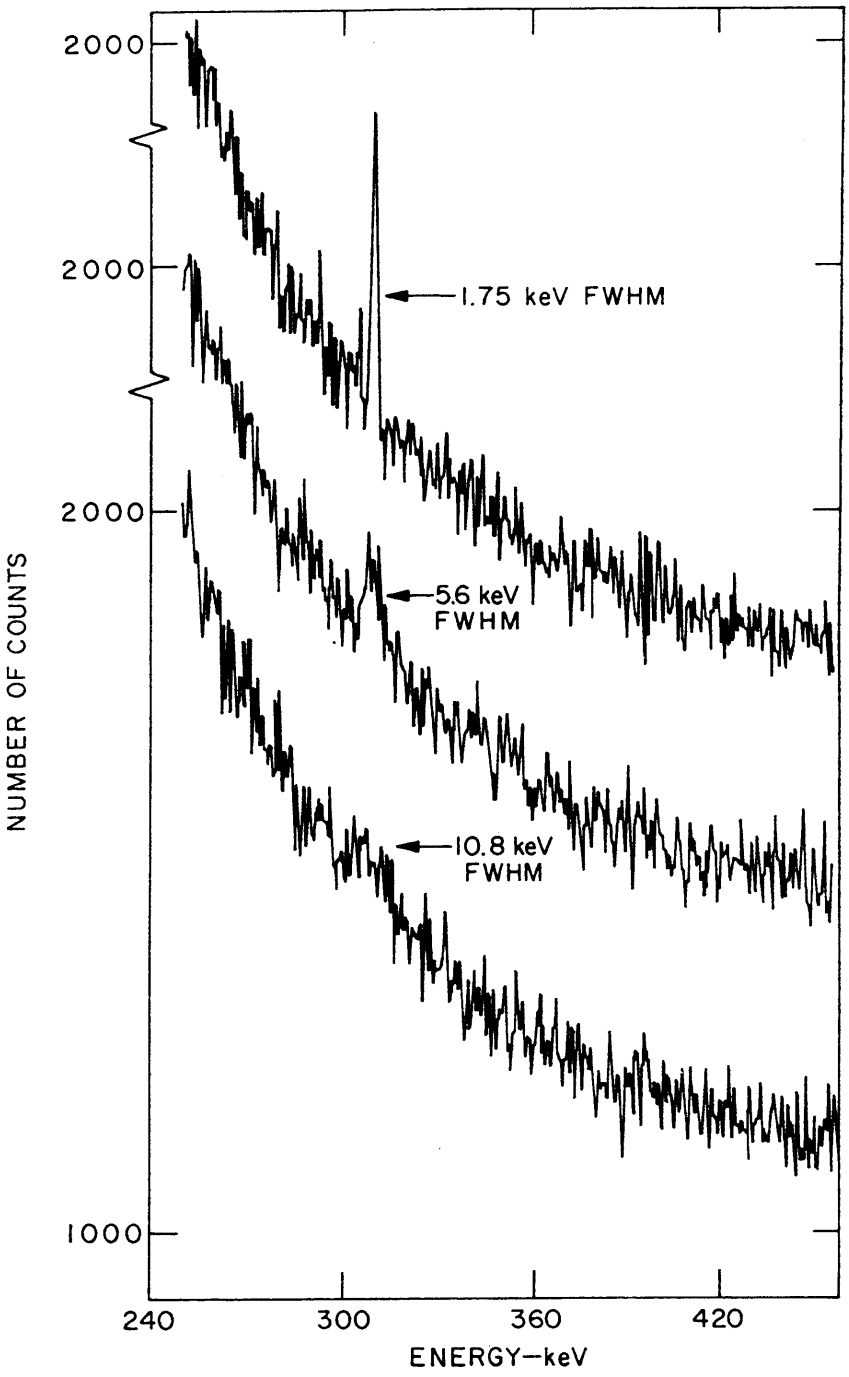
(J.Cl. Philippot, IEEE Trans. Nucl. Sci. **NS-17/3** (1970) 446)

NaI(Tl) scintillation detector: signal fluctuations

Ge detector: predominantly electronic noise

Resolution increases sensitivity

Signal to background ratio improves with better resolution
(narrow peak competes with fewer background counts)



G.A. Armantrout, *et al.*, IEEE Trans. Nucl. Sci. **NS-19/1** (1972) 107

Signal Fluctuations in a Scintillation Detector

Example: a typical NaI(Tl) system (from Derenzo)

511 keV gamma ray

↓

25000 photons in scintillator

↓

15000 photons at photocathode

↓

3000 photoelectrons at first dynode

↓

$3 \cdot 10^9$ electrons at anode

2 mA peak current

Resolution of energy measurement determined by statistical variance of produced signal quanta.

$$\frac{\Delta E}{E} = \frac{\Delta N}{N} = \frac{\sqrt{N}}{N} = \frac{1}{\sqrt{N}}$$

Resolution determined by smallest number of quanta in chain, i.e. number of photoelectrons arriving at first dynode.

In this example

$$\frac{\Delta E}{E} = \frac{1}{\sqrt{3000}} = 2\% \text{ r.m.s.} = 5\% \text{ FWHM}$$

Typically 7 – 8% obtained, due to non-uniformity of light collection and gain.

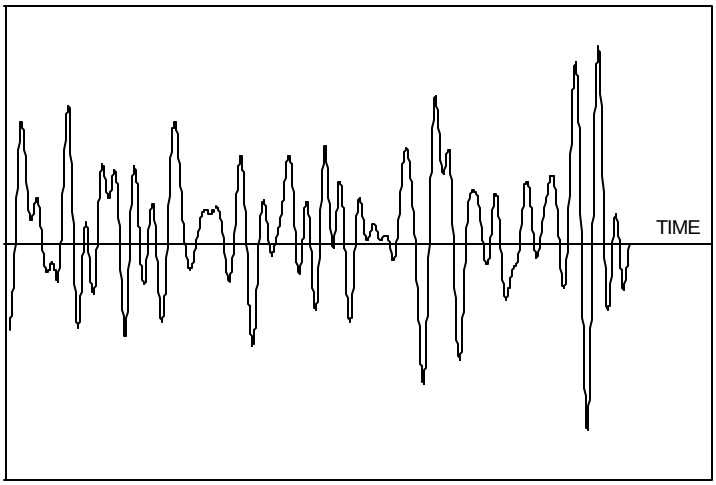
Baseline Fluctuations (Electronic Noise)

Choose a time when no signal is present.

Amplifier's quiescent output level (baseline):

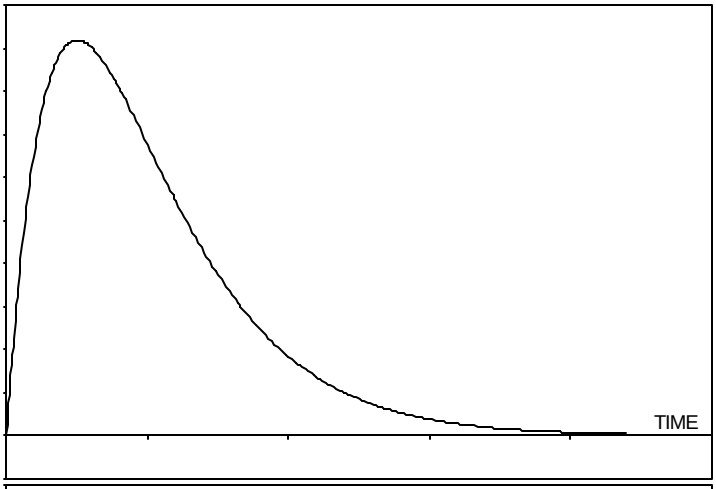
sensitivity x10

These fluctuations are added to any input signal

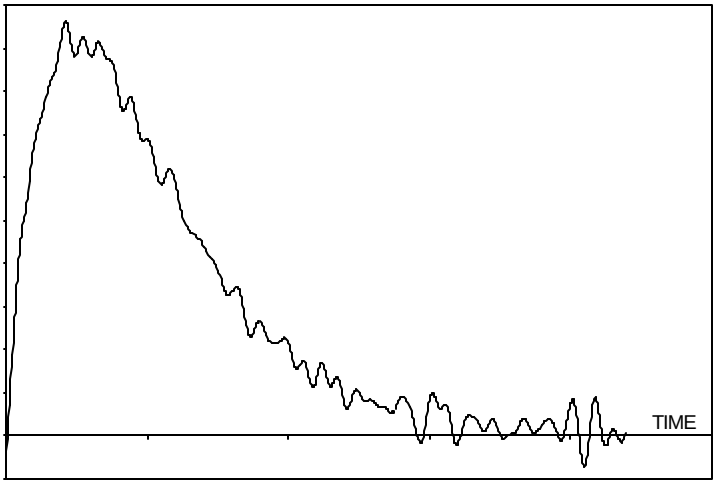


Pulse output of the ideal system

(sensitivity x1)



Signal + Noise



Measurement of peak amplitude yields
signal amplitude + noise fluctuation

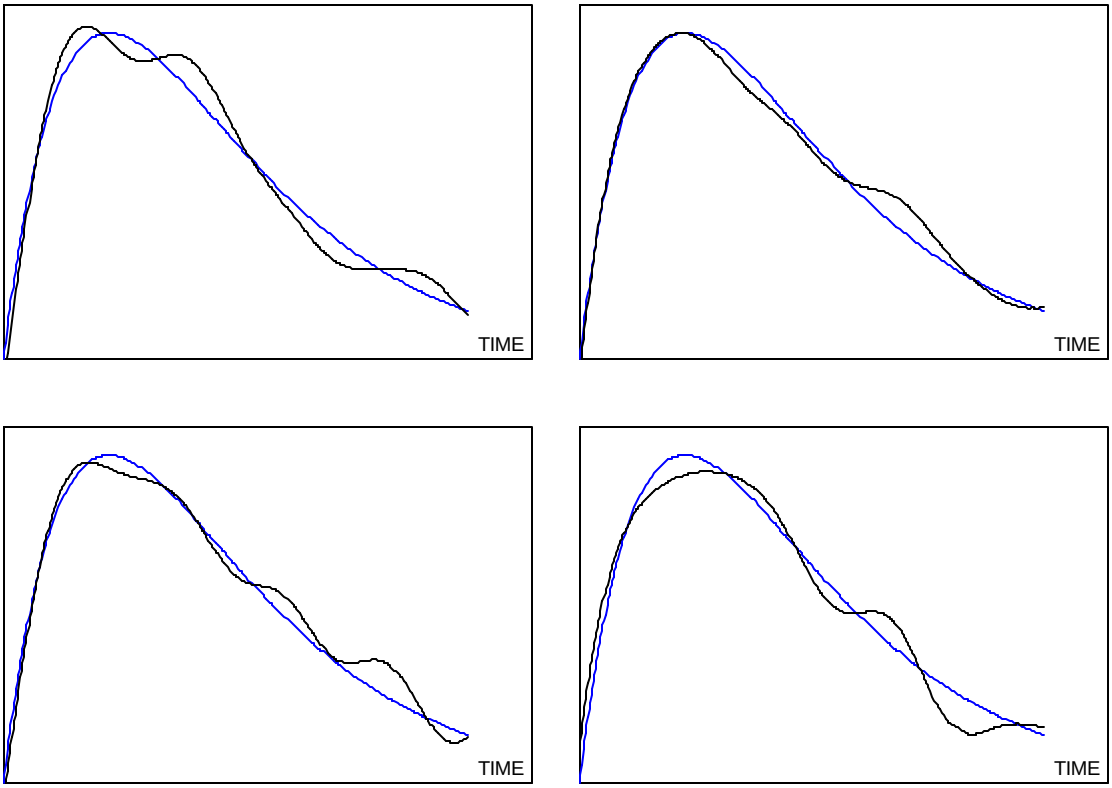
The preceding example could imply that the fluctuations tend to increase the measured amplitude, since the noise fluctuations vary more rapidly than the signal.

In an optimized system, the time scale of the fluctuation is comparable to the signal peaking time.

Then the measured amplitude fluctuates positive and negative relative to the ideal signal.

Measurements taken at 4 different times:

(noiseless signal superimposed for comparison)



Amplitude distribution of noise appears as amplitude distribution of signal.

3. The Problem

Radiation impinges on a sensor and creates an electrical signal.

The signal level is low and must be amplified to allow digitization and storage.

Both the sensor and amplifiers introduce signal fluctuations – noise.

1. Fluctuations in signal introduced by sensor

2. Noise from electronics superimposed on signal

The detection limit and measurement accuracy are determined by the signal-to-noise ratio.

Electronic noise affects all measurements:

1. Detect presence of hit:
   Noise level determines minimum threshold.
   If threshold too low, output dominated by noise hits.

2. Energy measurement:
   noise “smears” signal amplitude

3. Time measurement
   noise alters time dependence of signal pulse

How to optimize the signal-to-noise ratio?

1. Increase signal and reduce noise

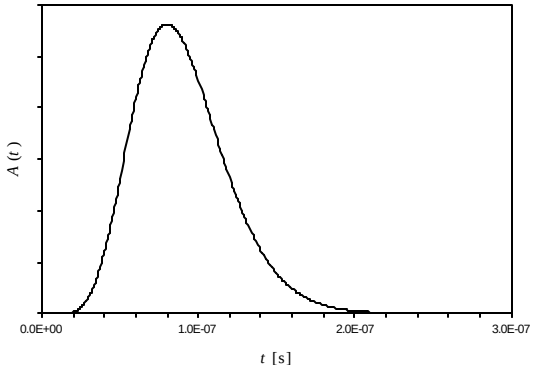
2. For a given sensor and signal: reduce electronic noise

Assume that the signal is a pulse.

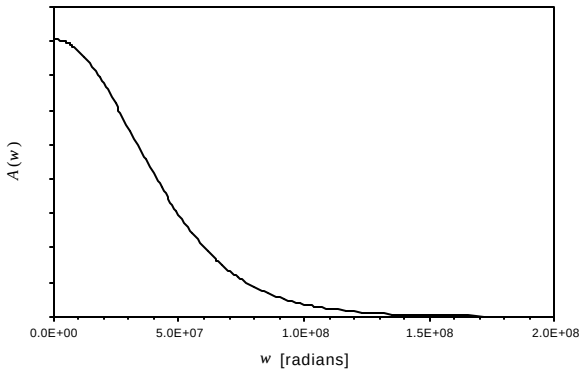
The time distribution of the signal corresponds to a frequency spectrum (Fourier transform).

Examples:

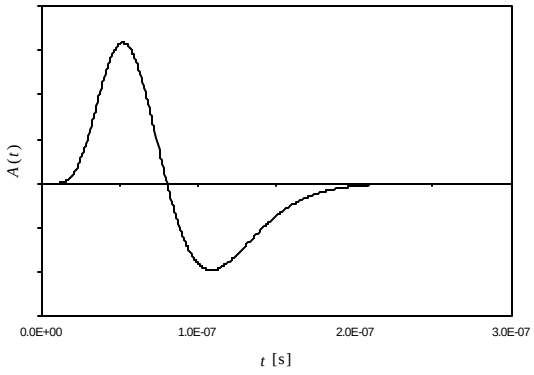
Time Domain

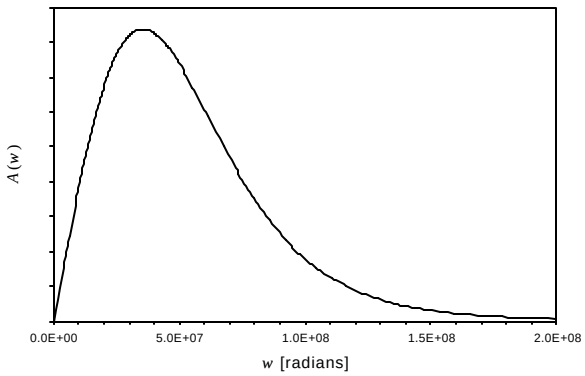


Frequency Domain



The pulse is unipolar, so it has a DC component and the frequency spectrum extends down to 0.

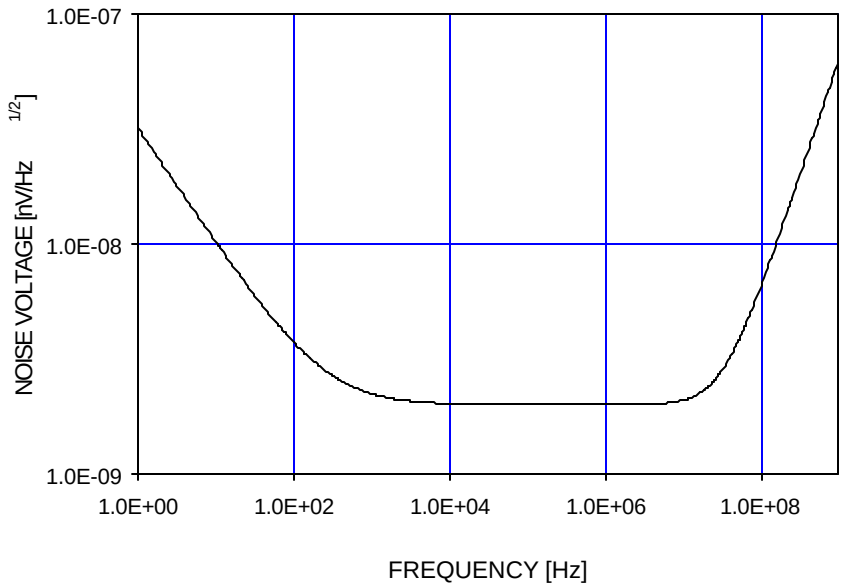




This bipolar pulse carries no net charge, so the frequency spectrum falls to zero at low frequencies.

The noise spectrum generally not the same as the signal spectrum.

Typical Noise Spectrum:



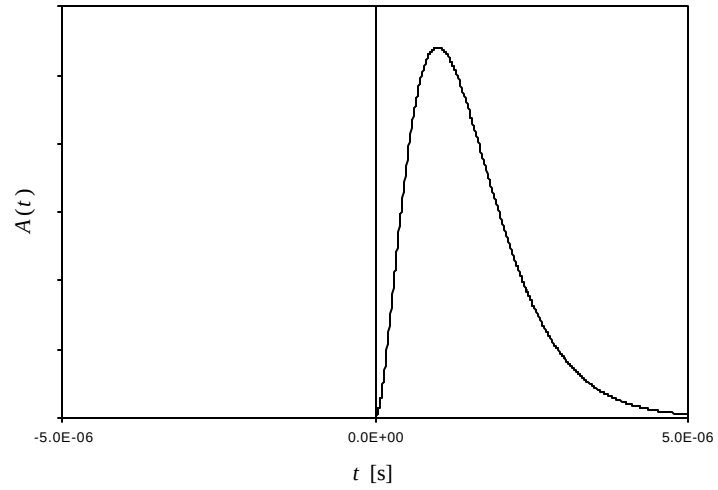
- tailor frequency response of measurement system to optimize signal-to-noise ratio.

Frequency response of measurement system affects both

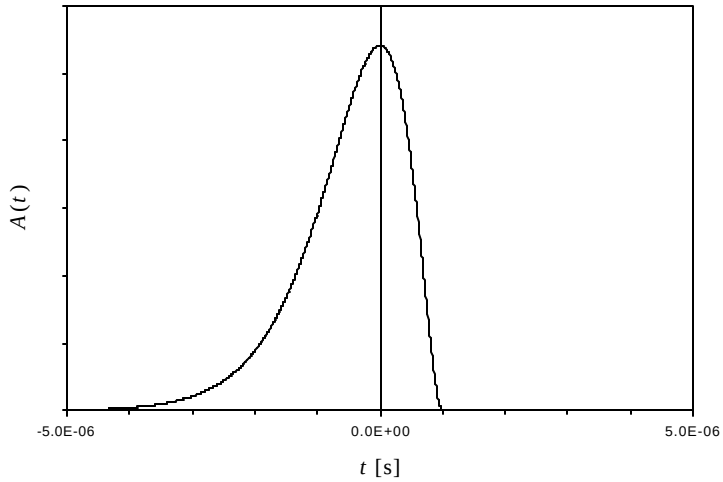
- signal amplitude and
- noise.

There is a general solution to this problem:
Apply a filter to make the noise spectrum white (constant over frequency). Then the optimum filter has an impulse response that is the signal pulse *mirrored in time* and shifted by the measurement time.

For example, if the signal pulse shape is:



The response of the optimum filter:



This is an “acausal” filter, i.e. it must act before the signal appears.

- only useful if the time of arrival is known in advance.

  Not good for random events

  – need time delay buffer memory → complexity!

Does that mean our problem is solved (and the lecture can end)?

1. The “optimum filter” preserves all information in signal, i.e. magnitude, timing, structure.

   Usually, we need only subset of the information content, i.e. area (charge) or time-of-arrival.

   Then the raw detector signal is not of the optimum form for the information that is required.

   For example, a short detector pulse would imply a fast filter function. This retains both amplitude and timing information. If only charge information is required, a slower filter is better, as will be shown later.

2. The optimum filter is often difficult or impractical to implement

   Digital signal processing would seem to remove this restriction, but this approach is not practical for very fast signals or systems that require low power.

4. Simpler filters often will do nearly as well

5. Even a digital system requires continuous (“analog”) pre-processing.

6. It’s often useful to understand what you’re doing, so we’ll spend some more time to bring out the physical background of signal formation and processing.

4. Signal processing systems

Large detector systems may consist of several subsystems especially designed to perform specific functions, for example

- position sensing (tracking)
- energy measurement (spectroscopy, calorimeters)
- timing
- particle identification

Functions

Although these subsystems may look very different and use radically differing technologies, they all tend to comprise the same basic functions:

1. Radiation deposits energy in a detecting medium.

   The medium may be gas, solid or liquid.

   In a tracking detector one wishes to detect the presence of a particle without affecting its trajectory, so the medium will be chosen to minimize energy loss and particle scattering.

   Conversely, if one wishes to measure the total energy (energy spectrometry or calorimetry), the absorber will be chosen to optimize energy loss (high density, high Z).

2. Energy is converted into an electrical signal, either directly or indirectly. Each detected particle will appear as a pulse of electric charge.

   Direct conversion:
   incident radiation ionizes atoms/molecules in absorber, creating mobile charges that are detected.
   (ionization chambers)

   Indirect conversion:
   incident radiation excites atomic/molecular states that decay by emission of light, which in a second step is converted into charge.
   (scintillation detectors)

   The primary signal charge is proportional to the energy absorbed.

   Some typical values of energy required to form a signal charge of 1 electron:

gases	30 eV
semiconductors	1 to 10 eV
scintillators	20 to 500 eV

   In neither of these schemes is the signal charge available instantaneously. In a scintillation detector the pulse duration is determined by the decay time of the optical transitions, in an ionization chamber the charges must move to the electrodes to obtain the full signal.

   Typical pulse durations: 1 ns – 10 μs

3. The electrical signal is amplified.

   a) electronic circuitry

   b) gain by secondary multiplication

      primary charge is accelerated to sufficient energy for it to liberate additional charge carriers by impact ionization.

      Examples: proportional chambers
      avalanche photodiodes
      photomultiplier

   Both techniques may introduce significant random fluctuations (electronic noise, avalanche noise).

   Ideally, a gain stage would increase only the magnitude of the detector pulse, without affecting its time dependence.

   This ideal behavior is never strictly realized in practice, as it would require amplifiers with infinite bandwidth.

   However, this is not a severe limitation, as in many applications it is quite acceptable and even desirable to change the pulse shape.

4. Pulse shaping

(not always necessary, but always present in some form)

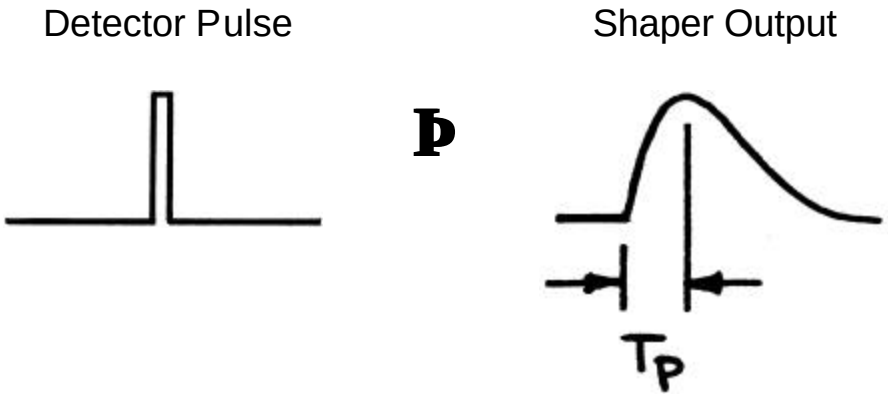
The time response of the system is tailored to optimize the measurement of signal magnitude or time and the rate of signal detection.

The output of the signal chain is a pulse (current or voltage) whose area is proportional to the original signal charge, i.e. the energy deposited in the detector.

Typically, the pulse shaper transforms a narrow detector current pulse to

- a broader pulse (to reduce electronic noise),
- with a gradually rounded maximum at the peaking time T_P

(to facilitate measurement of the amplitude)



However, to measure pulses in rapid succession, the duration of the pulse must be limited to avoid overlapping signals.

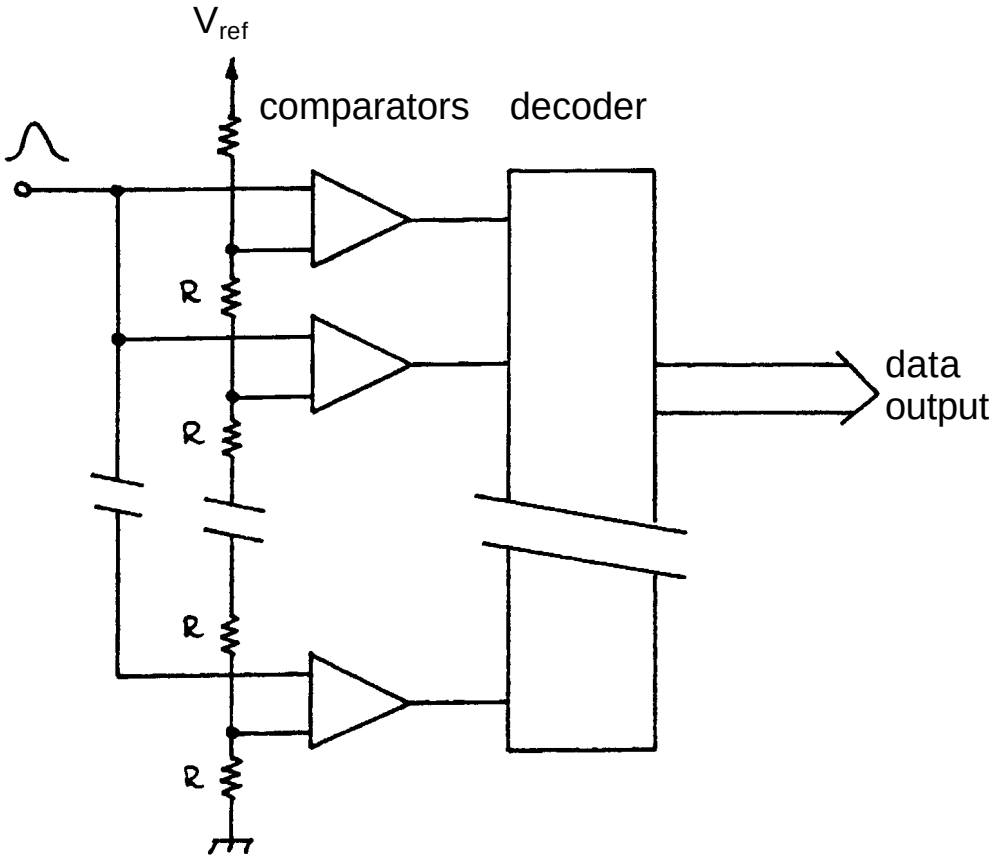
If the shape of the pulse does not change with signal level, the peak amplitude is also a measure of the energy, so one often speaks of pulse-height measurements or analysis.

The pulse height spectrum is the energy spectrum.

5. Digitization

a) signal magnitude
(analog-to-digital converter, viz. ADC or A/D)

Example:



The input signal is applied to n comparators in parallel. The switching thresholds are set by a resistor chain, such that the voltage difference between individual taps is equal to the desired measurement resolution.

In the presence of a signal all comparators with threshold levels less than the signal amplitude will fire. A decoder converts the parallel bit pattern into a more efficient form, for example binary code.

This type of ADC is fast, but requires as many comparators as measurement bins. Other converter types provide higher resolution and simpler circuitry at the expense of speed.

b) time difference between the detected signal and a reference signal
(time-to-digital converter, TDC)

The reference signal can be derived from another detector or from a common system clock, the crossing time of colliding beams, for example.

Circuit implementations include schemes that count “clock ticks” in fully digital circuitry or combine time-to-amplitude and amplitude-to-digital conversion in mixed analog-digital arrangements.

In complex detector systems the individual digitized outputs may require rather complex circuitry to combine the signal associated with a specific event and “package” them for efficient transfer.

2. Acquiring the Detector Signal

- Determine energy deposited in detector

- Detector signal generally a short current pulse

  Typical durations

Thin silicon detector (10 ... 300 μm thick):	100 ps – 30 ns
Thick (~cm) Si or Ge detector:	1 – 10 μs
Proportional chamber (gas):	10 ns – 10 μs
Gas microstrip or microgap chamber:	10 – 50 ns
Scintillator + PMT/APD:	100 ps – 10 μs

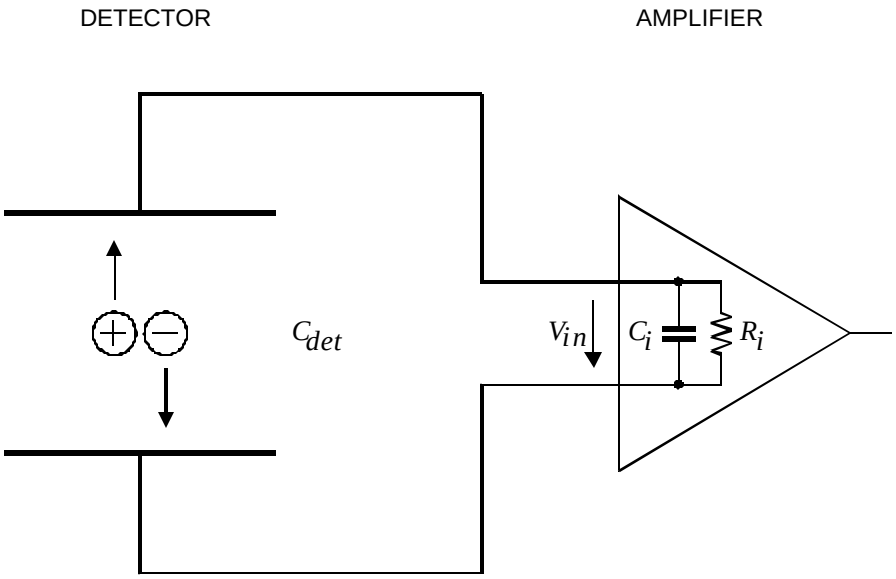
  The total charge Q_s contained in the detector current pulse $i_s(t)$ is proportional to the energy deposited in the detector

$$E \propto Q_s = \int i_s(t)\, dt$$

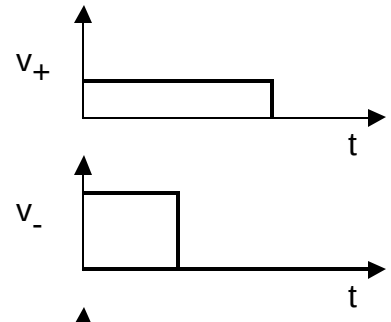
- Necessary to integrate the detector signal current.

  Possibilities:
  1. Integrate charge on input capacitance
  2. Use integrating (“charge sensitive”) preamplifier
  3. Amplify current pulse and use integrating (“charge sensing”) ADC

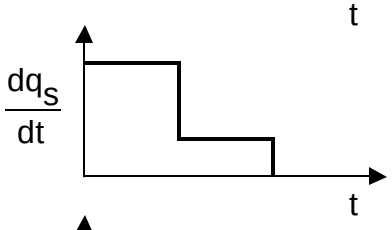
Integration on Input Capacitance



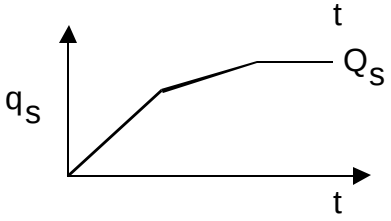
velocities of charge carriers



rate of induced charge on detector electrodes



signal charge



if R_i x (C_{det} + C_i) >> collection time

peak voltage at amplifier input

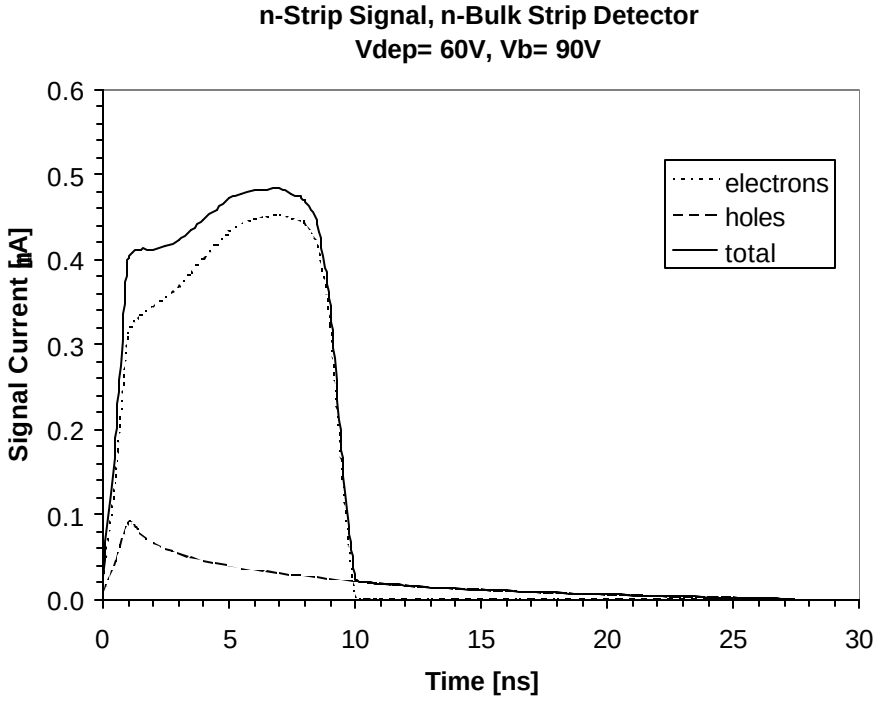
$$V_{in} = \frac{Q_s}{C_{\det} + C_i}$$

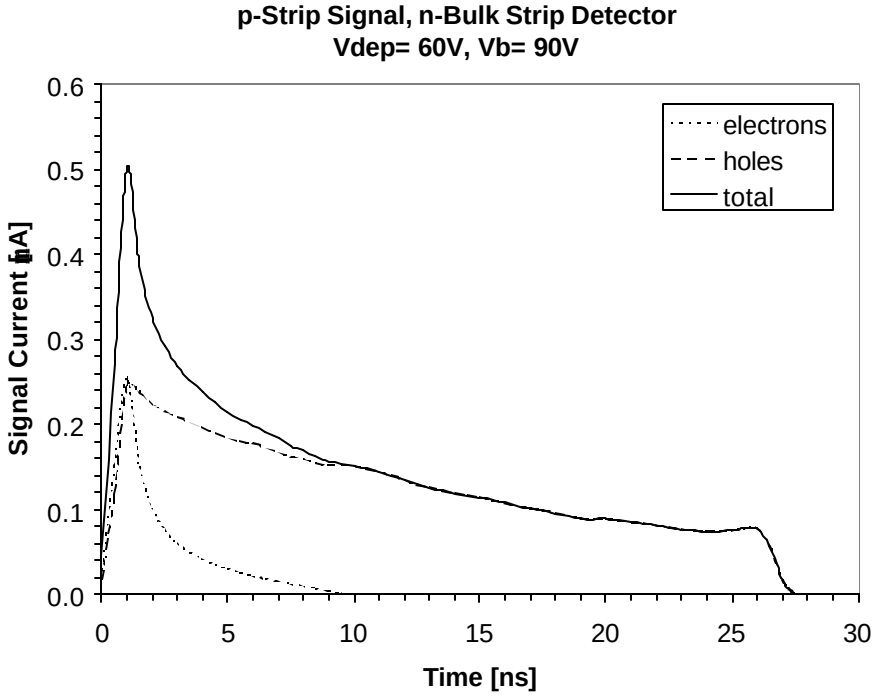
-

Magnitude of voltage depends on detector capacitance!

In reality the current pulses are more complex.

Current pulses on opposite sides (n-strip and p-strip) of a double-sided silicon strip detector (track traversing the detector)



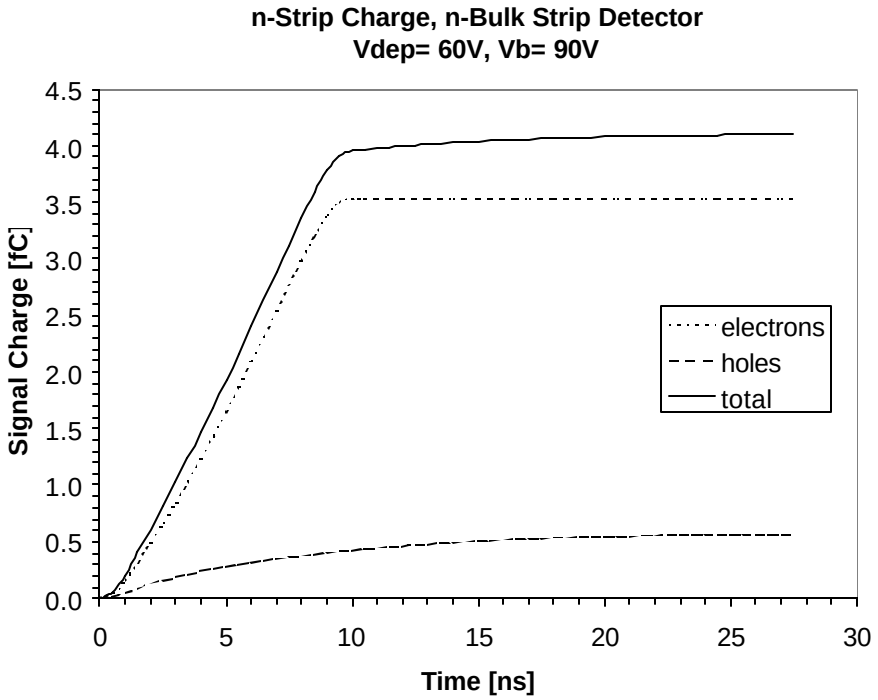


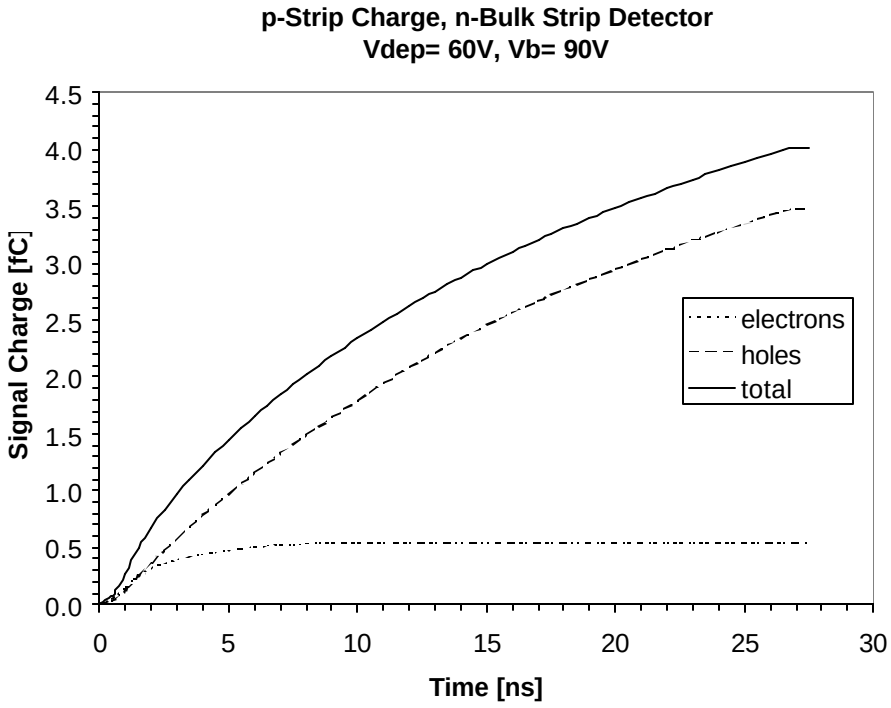
Although both pulses originate from the same particle track, the shapes are very different.

However, although the peak voltage or current signal measured by the amplifier may be quite different, the signal charge

$$Q_s = \int i_s dt$$

is the same.





Desirable to measure signal charge

- independent of detector pulse shape

When the input time constant RC is much greater than the signal duration, the peak voltage is a measure of the charge

$$V = \frac{1}{C}\int i_s dt = \frac{Q_s}{C}$$

The measured signal depends on the total capacitance at the input.

Awkward in system where the detector capacitance varies, e.g.

- different detector geometries
  (e.g. strip detectors with different lengths)
- varying detector capacitance
  (e.g. partially depleted semiconductor detectors)

Use system whose response is independent of detector capacitance.

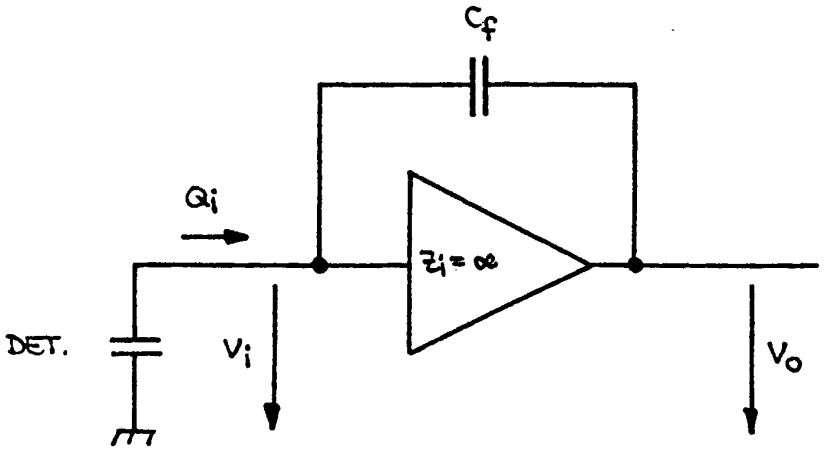
Active Integrator ("charge-sensitive amplifier")

Start with inverting voltage amplifier

Voltage gain $dv_o / dv_i = -A \quad \Rightarrow \quad v_o = -Av_i$

Input impedance = ∞ (i.e. no signal current flows into amplifier input)

Connect feedback capacitor C_f between output and input.



Voltage difference across C_f: $v_f = (A+1)v_i$

$\Rightarrow$ Charge deposited on C_f: $Q_f = C_f v_f = C_f (A+1) v_i$

$Q_i = Q_f$ (since $Z_i = \infty$)

$\Rightarrow$ Effective input capacitance

$$C_i = \frac{Q_i}{v_i} = C_f (A+1)$$

("dynamic" input capacitance)

Gain

$$A_Q = \frac{dV_o}{dQ_i} = \frac{A \cdot v_i}{C_i \cdot v_i} = \frac{A}{C_i} = \frac{A}{A+1} \cdot \frac{1}{C_f} \approx \frac{1}{C_f} \quad (A >> 1)$$

Q_i is the charge flowing into the preamplifier ....

but some charge remains on C_{det}.

What fraction of the signal charge is measured?

$$\frac{Q_i}{Q_s} = \frac{C_i v_i}{Q_{\det} + Q_i} = \frac{C_i}{Q_s} \cdot \frac{Q_s}{C_i + C_{\det}}$$

$$= \frac{1}{1 + \frac{C_{\det}}{C_i}} \approx 1 \ (\text{if } C_i >> C_{\det})$$

Example:

$A = 10^3$

C_f = 1 pF ⇒ C_i = 1 nF

C_{det} = 10 pF: Q_i / Q_s = 0.99

C_{det} = 500 pF: Q_i / Q_s = 0.67

↑

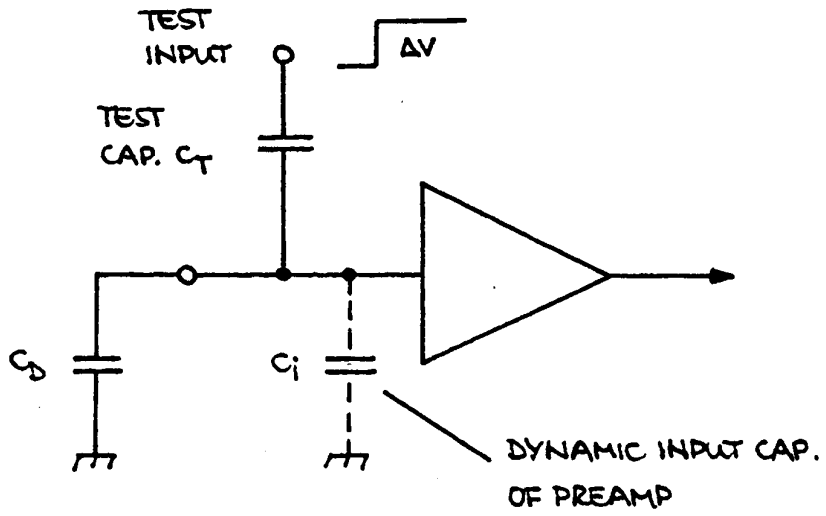
Si Det.: 50 μm thick
500 mm^2 area

Note: Input coupling capacitor must be >>C_i for high charge transfer efficiency.

Calibration

Inject specific quantity of charge - measure system response

Use voltage pulse (can be measured conveniently with oscilloscope)



$C_i >> C_T$ ⇒ Voltage step applied to test input develops over C_T.

⇒ $Q_T = \Delta V \cdot C_T$

Accurate expression:

$$Q_T = \frac{C_T}{1+\frac{C_T}{C_i}} \cdot \Delta V \approx C_T \left(1-\frac{C_T}{C_i}\right)\Delta V$$

Typically: $C_T/C_i = 10^{-3} - 10^{-4}$

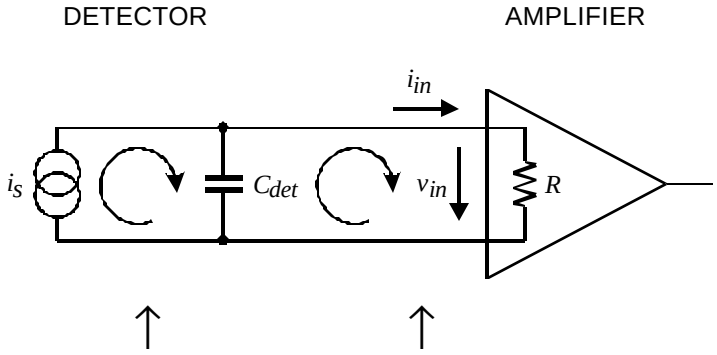
Realistic Charge-Sensitive Preamplifiers

The preceding discussion assumed idealized amplifiers with infinite speed.

In reality, amplifiers may be too slow to follow the instantaneous detector pulse.

Does this incur a loss of charge?

Equivalent Circuit:



charges moving in detector induce change of charge on detector electrodes

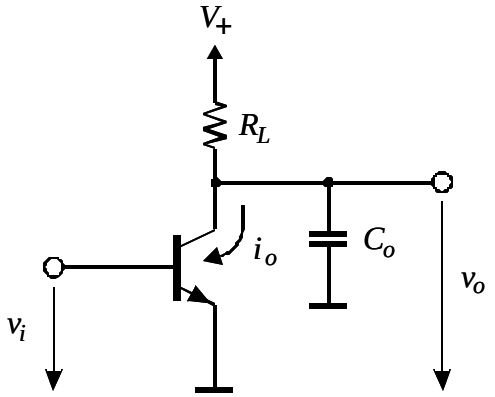
detector capacitance discharges into amplifier

Signal is preserved even if the amplifier responds much more slowly than the detector signal.

However, the response of the amplifier affects the measured pulse shape.

- How do “real” amplifiers affect the measured pulse shape?
- How does the detector affect amplifier response?

A Simple Amplifier



Voltage gain:

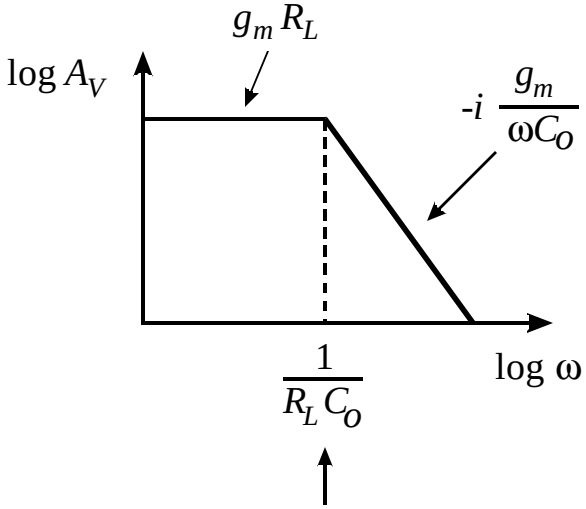
$$A_V = \frac{dv_o}{dv_i} = \frac{di_o}{dv_i} \cdot Z_L \equiv g_m Z_L$$

$g_m \equiv$ transconductance

$$Z_L = R_L \,//\, C_o$$

$$\frac{1}{Z_L} = \frac{1}{R_L} + \mathbf{i}w\ C_o \qquad \Rightarrow \qquad A_V = g_m \left(\frac{1}{R_L} + \mathbf{i}w\ C_o \right)^{-1}$$

low freq. high freq.



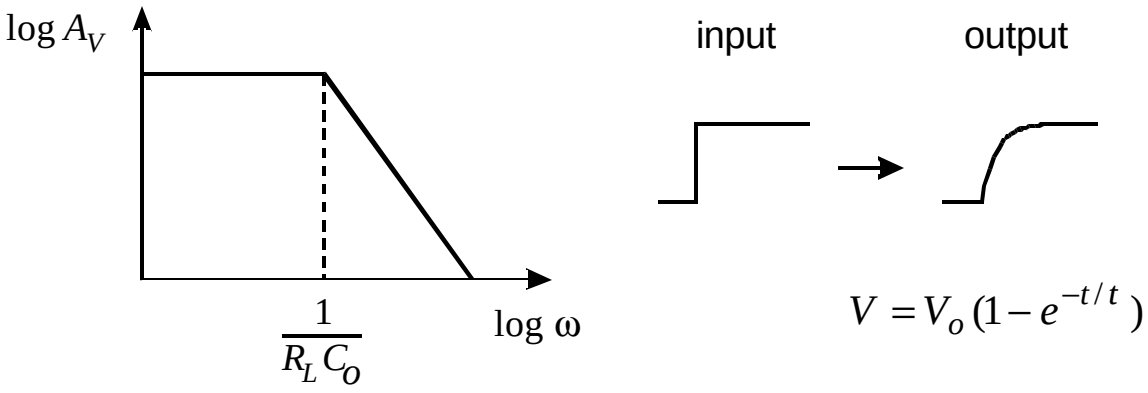
upper cutoff frequency $2\pi f_u$

Pulse Response of the Simple Amplifier

A voltage step $v_i(t)$ at the input causes a current step $i_o(t)$ at the output of the transistor.

For the output voltage to change, the output capacitance C_o must first charge up.

- The output voltage changes with a time constant $t = R_L C_o$



Frequency Domain Time Domain

The time constant t corresponds to the upper cutoff frequency

$$t = \frac{1}{2p\ f_u}$$

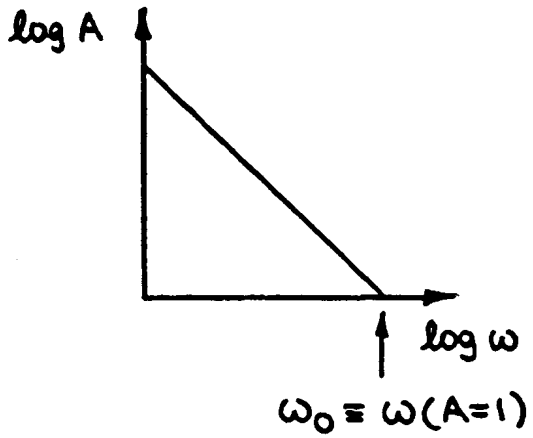
Input Impedance of a Charge-Sensitive Amplifier

Input impedance

$$Z_i = \frac{Z_f}{A+1} \approx \frac{Z_f}{A} \quad (A >> 1)$$

Amplifier gain vs. frequency beyond the upper cutoff frequency

$$A = -\mathbf{i}\ \frac{w_0}{w}$$



Feedback Impedance

$$Z_f = -\mathbf{i}\ \frac{1}{w\ C_f}$$

Gain-Bandwidth Product

$\Rightarrow$ Input Impedance

$$Z_i = -\frac{\mathbf{i}}{w\ C_f} \cdot \frac{1}{-\mathbf{i}\,\frac{w_0}{w}}$$

$$Z_i = \frac{1}{w_0 C_f}$$

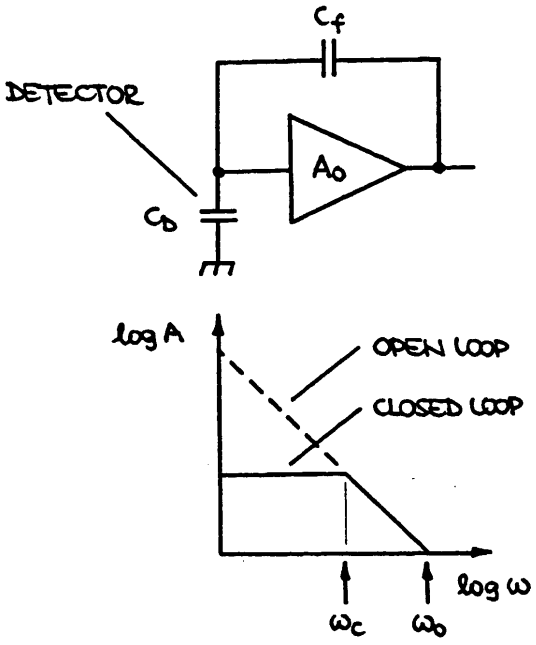
Imaginary component vanishes $\Rightarrow$ *Resistance:* $Z_i \rightarrow R_i$

$\Rightarrow$ low frequencies ($f < f_u$): capacitive input
high frequencies ($f > f_u$): resistive input

Time Response of a Charge-Sensitive Amplifier

Closed Loop Gain

$$A_f = \frac{C_D + C_f}{C_f} \quad (A_f << A_0)$$



$$A_f \approx \frac{C_D}{C_f} \quad (C_D >> C_f)$$

Closed Loop Bandwidth

$$w_C A_f = w_0$$

Response Time

$$t_{amp} = \frac{1}{w_C} = C_D \frac{1}{w_0 C_f}$$

- **Rise time increases with detector capacitance.**

Alternative Picture: Input Time Constant

$$t_i = R_i C_D$$

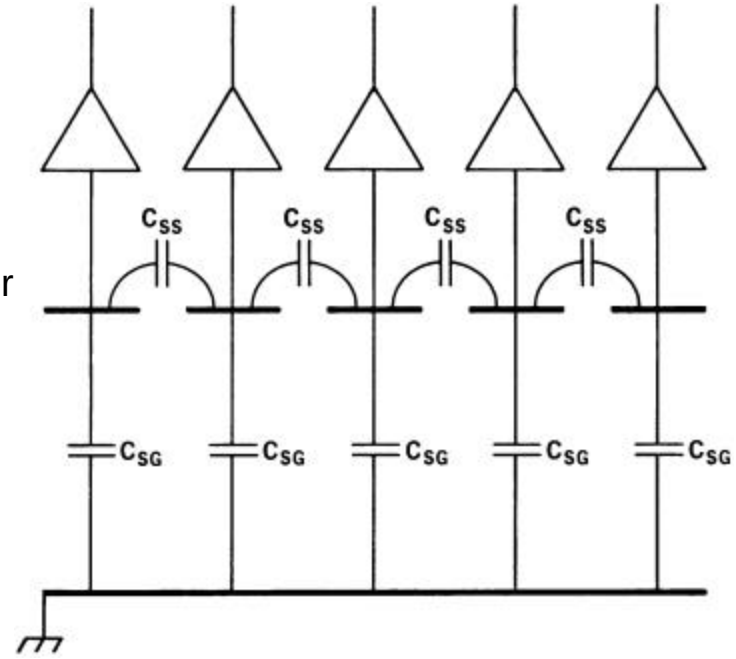
$$t_i = \frac{1}{w_0 C_f} \cdot C_D = t_{amp}$$

Same result as from conventional feedback theory.

Input impedance is critical in strip or pixel detectors:

Amplifiers must have a low input impedance to reduce transfer of charge through capacitance to neighboring strips

strip detector electrodes



For strip pitches that are smaller than the bulk thickness the capacitance is dominated by the fringing capacitance to the neighboring strips C_{SS}.

Typically: 1 – 2 pF/cm for strip pitches of 25 – 100 µm on Si.

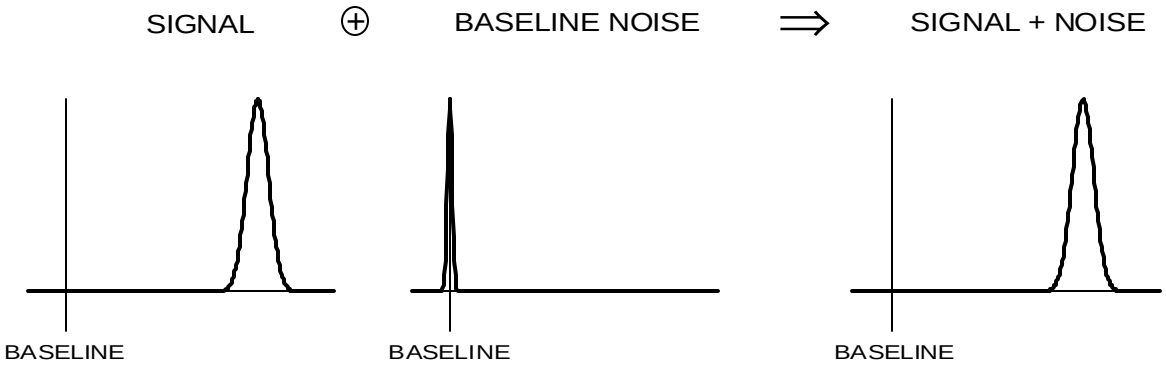
The backplane capacitance C_{SG} is typically 20% of the strip-to-strip capacitance.

Negligible cross-coupling at times t > (2 … 3) x $R_i C_D$ and if $C_i >> C_D$.

3. Resolution and Signal-to-Noise Ratio

What determines Resolution?

1. Signal Variance >> Baseline Variance



- Electronic (baseline) noise not important

  Examples:

  - High-gain proportional chambers
  - Scintillation Counters with High-Gain PMTs

    e.g. 1 MeV γ-rays absorbed by NaI(Tl) crystal

    Number of photoelectrons

    $N_{pe} \approx 8 \cdot 10^4$ [MeV^{-1}] $\times E_g \times QE \approx 2.4 \cdot 10^4$

    Variance typically

    $s_{pe} = N_{pe}^{1/2} \approx 160$ and $s_{pe} / N_{pe} \approx 5$ - 8%

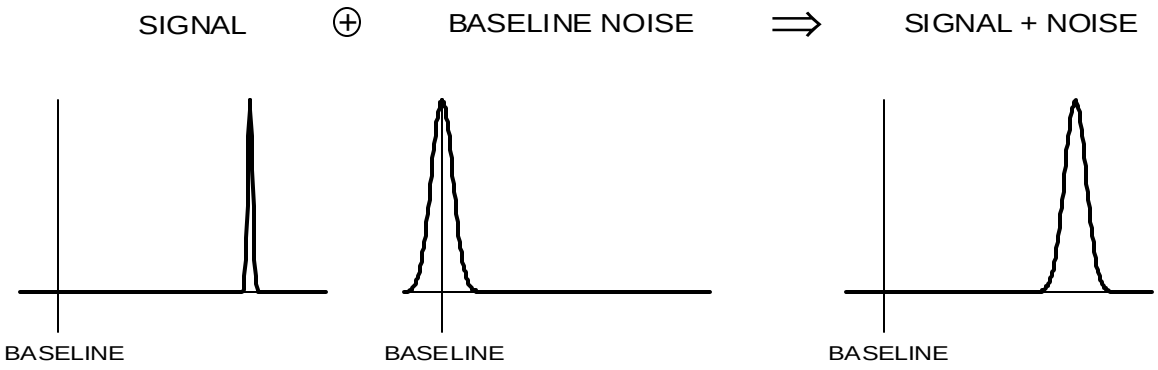
    Signal at PMT anode (assume Gain= 10^4)

    $Q_{sig} = G_{PMT} N_{pe} \approx 2.4 \cdot 10^8$ el and

    $s_{sig} = G_{PMT} s_{pe} \approx 1.2 \cdot 10^7$ el

    whereas electronic noise easily < 10^4 el

2. Signal Variance << Baseline Variance



- Electronic (baseline) noise critical for resolution

  Examples

  - Gaseous ionization chambers (no internal gain)
  - Semiconductor detectors

  e.g. in Si

  Number of electron-hole pairs

  $$N_{ep} = E_{dep} / (3.6 \text{ eV})$$

  Variance

  $$\sigma_{ep} = \sqrt{F \cdot N_{ep}}$$

  (where F = Fano factor ≈ 0.1)

  For 50 keV photons

  $$\sigma_{ep} \approx 40 \text{ el} \Rightarrow \quad \sigma_{ep} / N_{ep} = 7.5 \cdot 10^{-4}$$

  obtainable noise levels are 10 to 1000 el.

Baseline fluctuations can have many origins ...

pickup of external interference

artifacts due to imperfect electronics

… etc.,

but the (practical) fundamental limit is electronic noise.

Basic Noise Mechanisms

Consider n carriers of charge e moving with a velocity v through a sample of length l. The induced current i at the ends of the sample is

$$i = \frac{n\,e\,v}{l}.$$

The fluctuation of this current is given by the total differential

$$\langle di \rangle^2 = \left(\frac{ne}{l} \langle dv \rangle \right)^2 + \left(\frac{ev}{l} \langle dn \rangle \right)^2$$

where the two terms are added in quadrature since they are statistically uncorrelated.

Two mechanisms contribute to the total noise:

- velocity fluctuations, *e.g.* thermal noise
- number fluctuations, *e.g.* shot noise
  excess or '1/ f ' noise

Thermal noise and shot noise are both “white” noise sources, i.e.

power per unit bandwidth is constant:

≡ spectral density)

$$\frac{dP_{noise}}{df} = const.$$

or

$$\frac{dV_{noise}^2}{df} = const. \equiv e_n^2$$

whereas for “1/ f ” noise

$$\frac{dP_{noise}}{df} = \frac{1}{f^a}$$

(typically a = 0.5 – 2)

1. Thermal Noise in Resistors

The most common example of noise due to velocity fluctuations is the thermal noise of resistors.

Spectral noise power density vs. frequency f

$$\frac{dP_{noise}}{df} = 4kT$$

k = Boltzmann constant
T = absolute temperature

since

$$P = \frac{V^2}{R} = I^2 R$$

R = DC resistance

the spectral noise voltage density

$$\frac{dV_{noise}^2}{df} \equiv e_n^2 = 4kTR$$

and the spectral noise current density

$$\frac{dI_{noise}^2}{df} \equiv i_n^2 = \frac{4kT}{R}$$

The total noise depends on the bandwidth of the system.
For example, the total noise voltage at the output of a voltage amplifier with the frequency dependent gain $A_v(f)$ is

$$v_{on}^2 = \int_0^\infty e_n^2 \; A_v^2(f) \; df$$

Note: Since spectral noise components are non-correlated, one must integrate over the noise power.

2. Shot noise

A common example of noise due to number fluctuations is "shot noise", which occurs whenever carriers are injected into a sample volume independently of one another.

Example: current flow in a semiconductor diode (emission over a barrier)

Spectral noise current density:

$$i_n^2 = 2q_e I$$

q_e = electron charge
I = DC current

A more intuitive interpretation of this expression will be given later.

Note: Shot noise does not occur in "ohmic" conductors. Since the number of available charges is not limited, the fields caused by local fluctuations in the charge density draw in additional carriers to equalize the total number.

- For derivations of the thermal and shot noise spectral densities, see Appendix 1.

Noise Bandwidth vs. Signal Bandwidth

Consider an amplifier with the frequency response *A(f)*. This can be rewritten

$$A(f) \equiv A_0 G(f)$$

where A_0 is the maximum gain and $G(f)$ describes the frequency response.
For example, in the simple amplifier described above

$$A_V = g_m \left(\frac{1}{R_L} + \mathbf{i}w\ C_o \right)^{-1} = g_m R_L \ \frac{1}{1 + \mathbf{i}w R_L C_o}$$

and using the above convention

$$A_0 \equiv g_m R_L \quad \text{and} \quad G(f) \equiv \frac{1}{1 + \mathbf{i}\ (2\pi f\ R_L C_o)}$$

If a “white” noise source with spectral density e_{ni} is present at the input, the total noise voltage at the output is

$$V_{no} = \sqrt{\int_0^\infty e_{ni}^2 \left| A_0 G(f) \right|^2 df} = e_{ni}\ A_0 \sqrt{\int_0^\infty G^2(f) df} \equiv e_{ni}\ A_0 \sqrt{\Delta f_n}$$

Δf_n is the “noise bandwidth”.

Note that, in general, the noise bandwidth and the signal bandwidth are not the same. If the upper cutoff frequency is determined by a single RC time constant, as in the “simple amplifier”, the signal bandwidth

$$\Delta f_s = f_u = \frac{1}{2\pi\ RC}$$

and the noise bandwidth

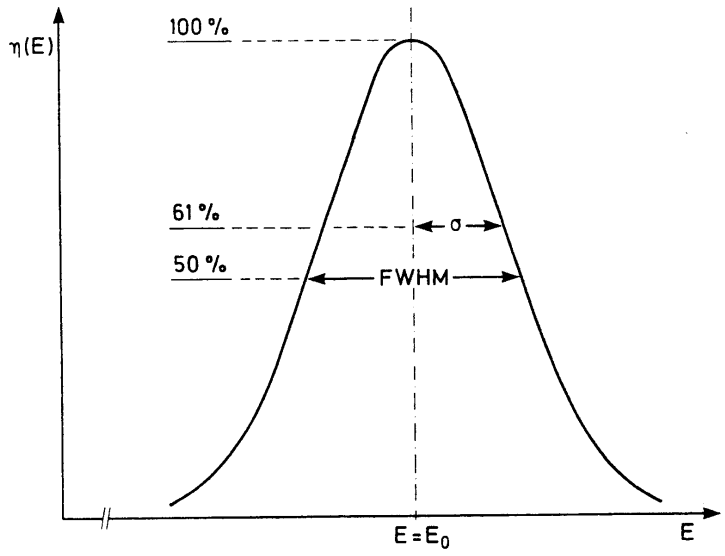
$$\Delta f_n = \frac{1}{4RC} = \frac{\pi}{2}\ f_u$$

Independent noise contributions add in quadrature
(additive in noise power)

$$V_{n,tot} = \sqrt{\sum_i V_{ni}^2}$$

Both thermal and shot noise are purely random.

- amplitude distribution is gaussian



- noise modulates baseline
- baseline fluctuations superimposed on signal
- output signal has gaussian distribution

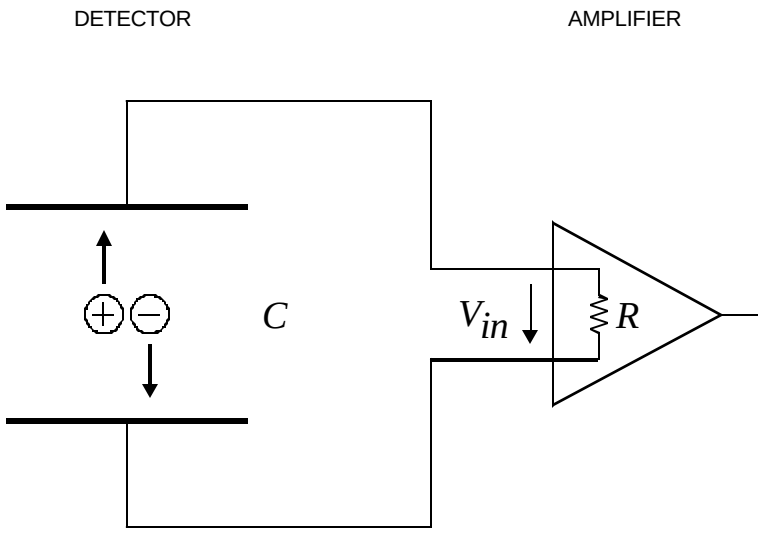
Measuring Resolution

Inject an input signal with known charge using a pulse generator set to approximate the detector signal (possible ballistic deficit). Measure the pulse height spectrum.

peak centroid → signal magnitude

peak width → noise (FWHM= 2.35 rms)

Signal-to-Noise Ratio vs. Detector Capacitance



At long input time constants $t = RC$ the detector signal current is integrated on the detector capacitance.

The resulting voltage sensed by the amplifier

$$V_{in} = \frac{Q_{\text{det}}}{C} = \frac{\int i_s dt}{C}$$

Then the peak amplifier signal is inversely proportional to the **total capacitance at the input**, i.e. the sum of

detector capacitance,
input capacitance of the amplifier, and
stray capacitances.

For a constant noise voltage v_n , the signal-to-noise ratio

$$\frac{S}{N} = \frac{V_S}{v_n} \propto \frac{1}{C}$$

- However, S/N does not become infinite as $C \to 0$ (see Appendix 2)

Charge-Sensitive Preamplifier
Noise vs. Detector Capacitance

In a voltage-sensitive preamplifier

- noise voltage at the output is essentially independent of detector capacitance,

  i.e. the *equivalent input noise voltage* $v_{ni} = v_{no}/A_v$.

- input signal decreases with increasing input capacitance, so signal-to-noise ratio depends on detector capacitance.

In a charge-sensitive preamplifier, the signal at the amplifier output is independent of detector capacitance (if $C_i >> C_{det}$).

What is the noise behavior?

- Noise appearing at the output of the preamplifier is fed back to the input, decreasing the output noise from the open-loop value $v_{no} = v_{ni} A_{v0}$.

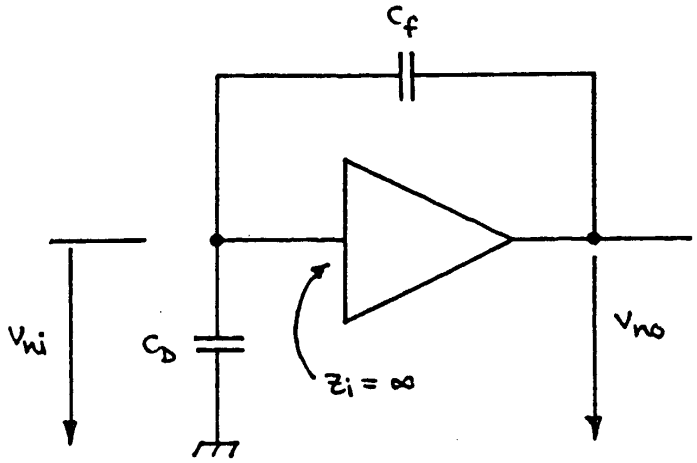
- The magnitude of the feedback depends on the shunt impedance at the input, i.e. the detector capacitance.

Note, that although specified as an equivalent input noise, the dominant noise sources are typically internal to the amplifier. Only in a fed-back configuration is some of this noise actually present at the input. In other words, the primary noise signal is not a physical charge (or voltage) at the amplifier input, to which the loop responds in the same manner as to a detector signal.

- **S/N at the amplifier output depends on amount of feedback.**

Noise in charge-sensitive preamplifiers

Start with an output noise voltage v_{no}, which is fed back to the input through the capacitive voltage divider $C_f - C_d$.



$$v_{no} = v_{ni}\frac{X_{C_f} + X_{C_D}}{X_{C_D}} = v_{ni}\frac{\dfrac{1}{w\ C_f} + \dfrac{1}{w\ C_D}}{\dfrac{1}{w\ C_D}}$$

$$v_{no} = v_{ni}\left(1 + \frac{C_D}{C_f}\right)$$

Equivalent input noise charge

$$Q_{ni} = \frac{v_{no}}{A_Q} = v_{no}C_f$$

$$Q_{ni} = v_{ni}\left(C_D + C_f\right)$$

Signal-to-noise ratio

$$\frac{Q_s}{Q_{ni}} = \frac{Q_s}{v_{ni}(C_D + C_f)} = \frac{1}{C}\ \frac{Q_s}{v_{ni}}$$

Same result as for voltage-sensitive amplifier, but here

- *the signal is constant and*
- *the noise grows with increasing C.*

As shown previously, the pulse rise time at the amplifier output also increases with total capacitive input load C, because of reduced feedback.

In contrast, the rise time of a voltage sensitive amplifier is not affected by the input capacitance, although the equivalent noise charge increases with C just as for the charge-sensitive amplifier.

Conclusion

In general

- optimum S/N is independent of whether the voltage, current, or charge signal is sensed.
- S/N cannot be *improved* by feedback.

Practical considerations, i.e. type of detector, amplifier technology, can favor one configuration over the other.

Pulse Shaping

Two conflicting objectives:

1. Improve Signal-to-Noise Ratio S/N

   Restrict bandwidth to match measurement time

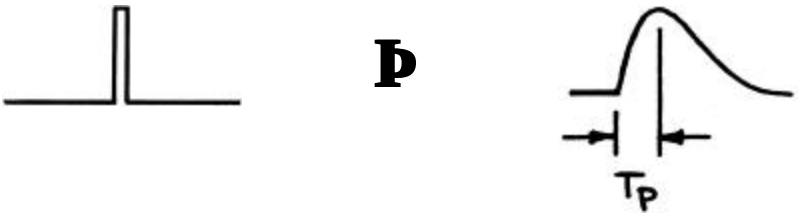
   ⇒ Increase pulse width

   Typically, the pulse shaper transforms a narrow detector current pulse to

   a broader pulse
   (to reduce electronic noise),

   with a gradually rounded maximum at the peaking time T_P
   (to facilitate measurement of the amplitude)

Detector Pulse ⇒ Shaper Output

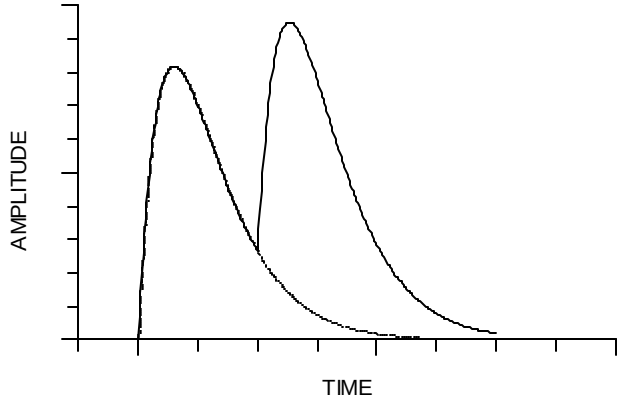


If the shape of the pulse does not change with signal level, the peak amplitude is also a measure of the energy, so one often speaks of pulse-height measurements or pulse height analysis. The pulse height spectrum is the energy spectrum.

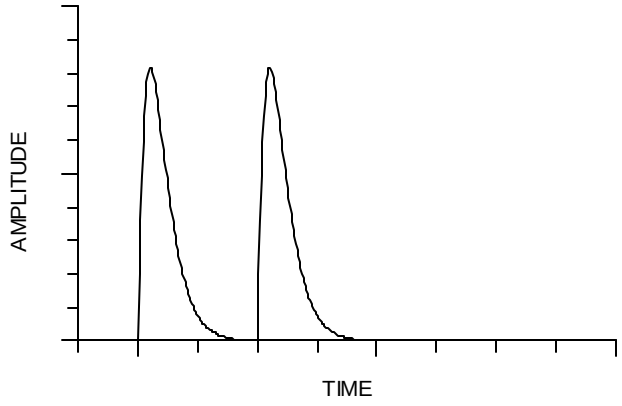
2. Improve Pulse Pair Resolution

- Decrease pulse width

Pulse pile-up distorts amplitude measurement



Reducing pulse shaping time to 1/3 eliminates pile-up.



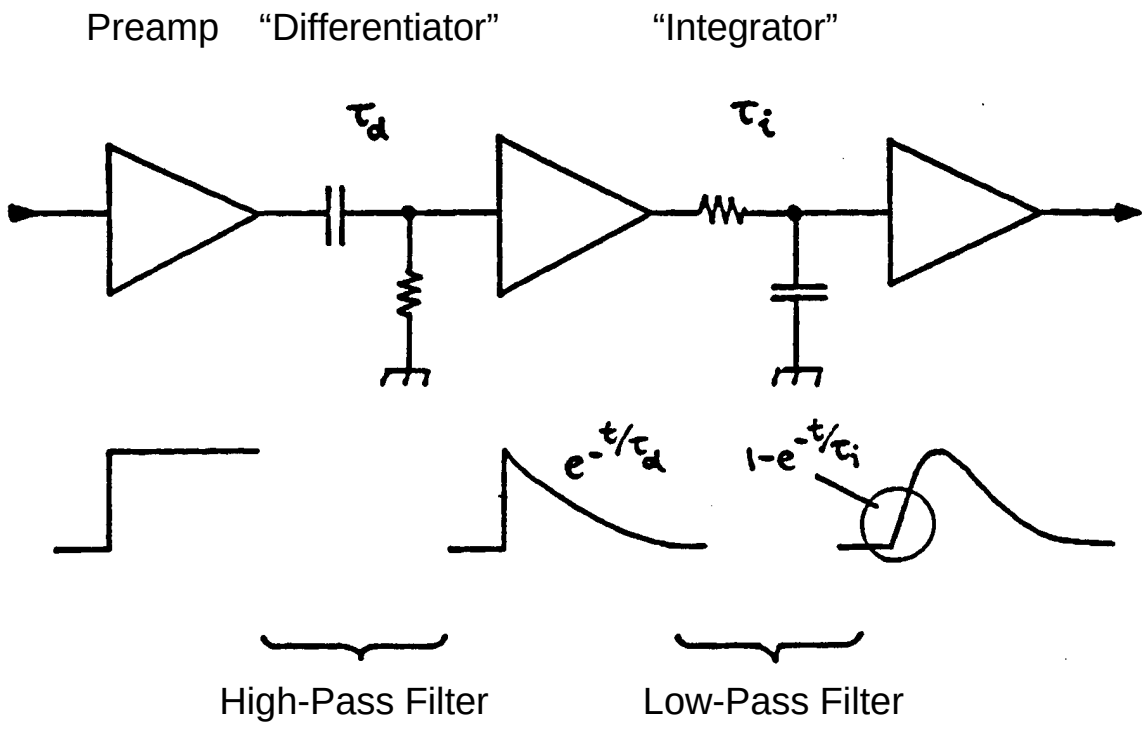
Necessary to find balance between these conflicting requirements. Sometimes minimum noise is crucial, sometimes rate capability is paramount.

Usually, many considerations combined lead to a "non-textbook" compromise.

- *"Optimum shaping" depends on the application!*

- Shapers need not be complicated –
  Every amplifier is a pulse shaper!

Simple Example: CR-RC Shaping



Simple arrangement: Noise performance only 36% worse than optimum filter with same time constants.

Þ Useful for estimates, since simple to evaluate

Key elements

- lower frequency bound
- upper frequency bound
- signal attenuation

important in all shapers.

Pulse Shaping and Signal-to-Noise Ratio

Pulse shaping affects both the

- total noise

and

- peak signal amplitude

at the output of the shaper.

Equivalent Noise Charge

Inject known signal charge into preamp input
(either via test input or known energy in detector).

Determine signal-to-noise ratio at shaper output.

Equivalent Noise Charge $\equiv$ Input charge for which $S/N = 1$

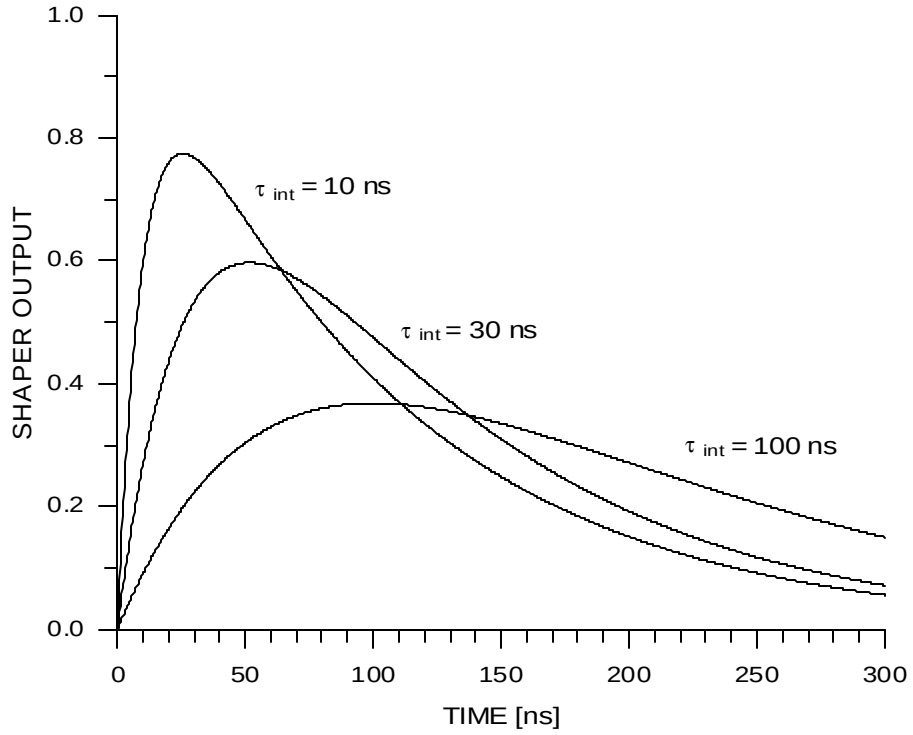
Effect of relative time constants

Consider a *CR-RC* shaper with a fixed differentiator time constant of 100 ns.

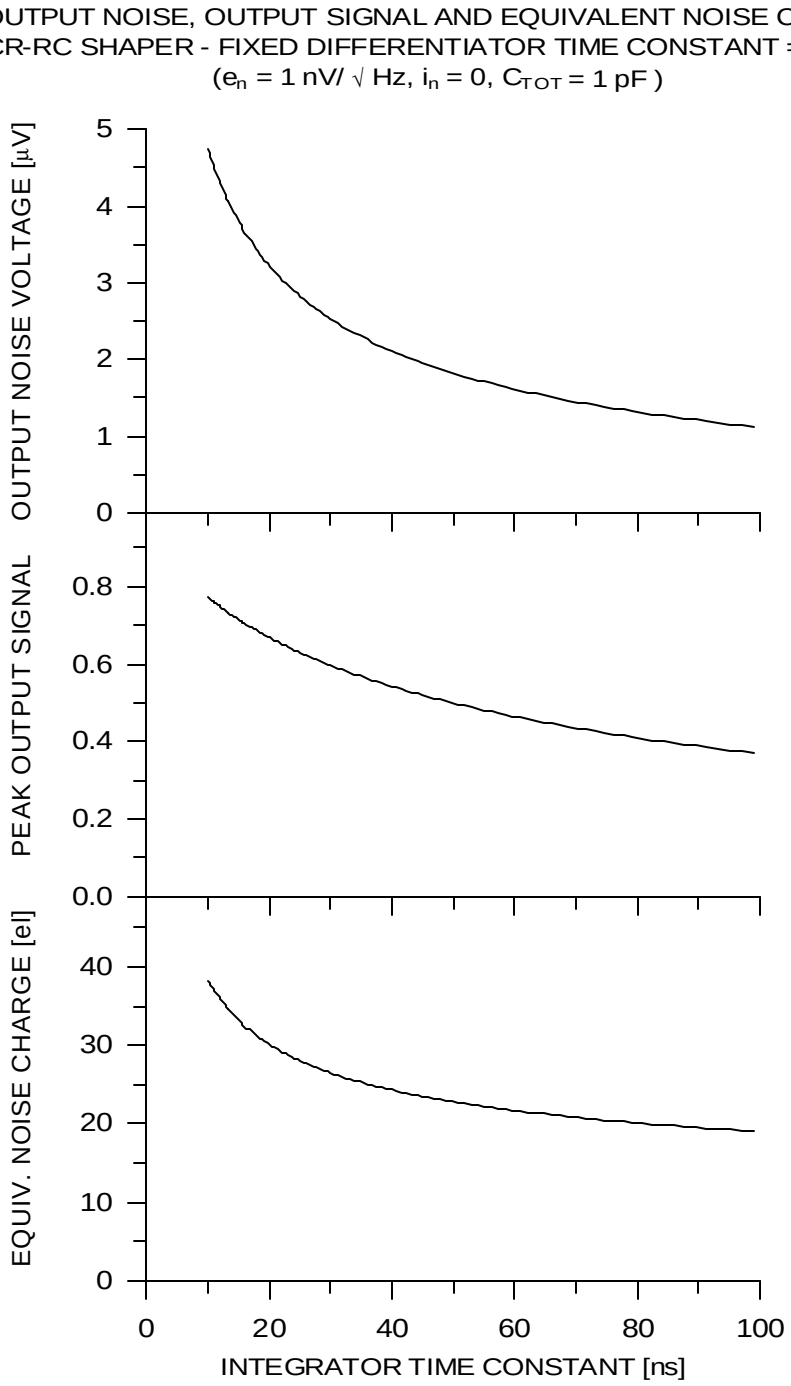
Increasing the integrator time constant lowers the upper cut-off frequency, which decreases the total noise at the shaper output.

However, the peak signal also decreases.

CR-RC SHAPER
FIXED DIFFERENTIATOR TIME CONSTANT = 100 ns
INTEGRATOR TIME CONSTANT = 10, 30 and 100 ns



Still keeping the differentiator time constant fixed at 100 ns, the next set of graphs shows the variation of output noise and peak signal as the integrator time constant is increased from 10 to 100 ns.



$$\frac{V_{no}(100\,\text{ns})}{V_{no}(10\,\text{ns})} = \frac{1}{4.2}$$

$$\frac{V_{so}(100\,\text{ns})}{V_{so}(10\,\text{ns})} = \frac{1}{2.1}$$

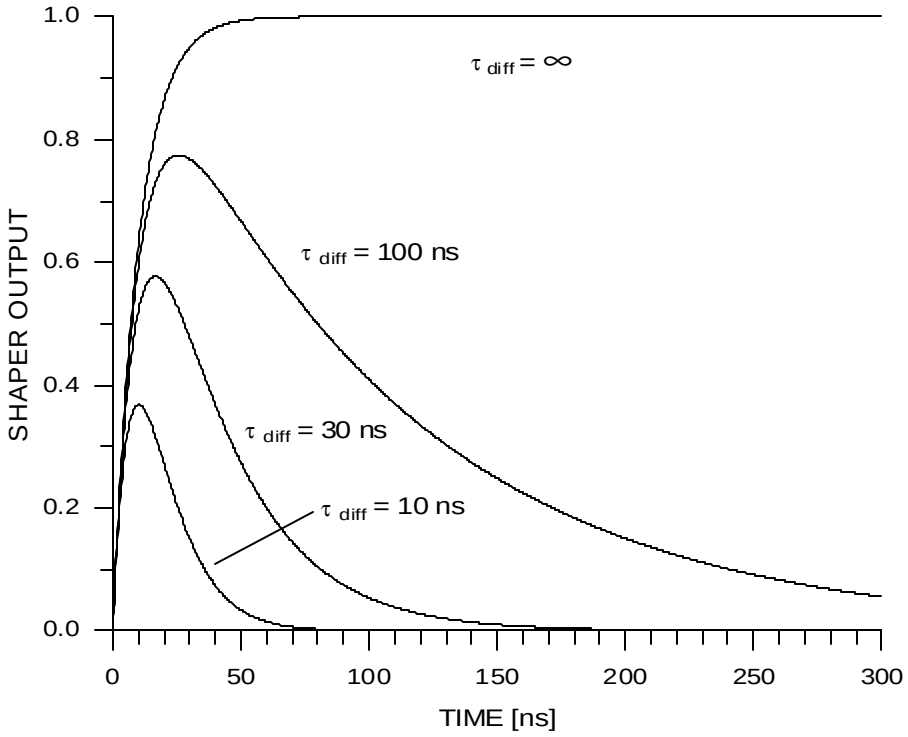
The roughly 4-fold decrease in noise is partially compensated by the 2-fold reduction in signal, so that

$$\frac{Q_n(100\,\text{ns})}{Q_n(10\,\text{ns})} = \frac{1}{2}$$

For comparison, consider the same *CR-RC* shaper with the integrator time constant fixed at 10 ns and the differentiator time constant variable.

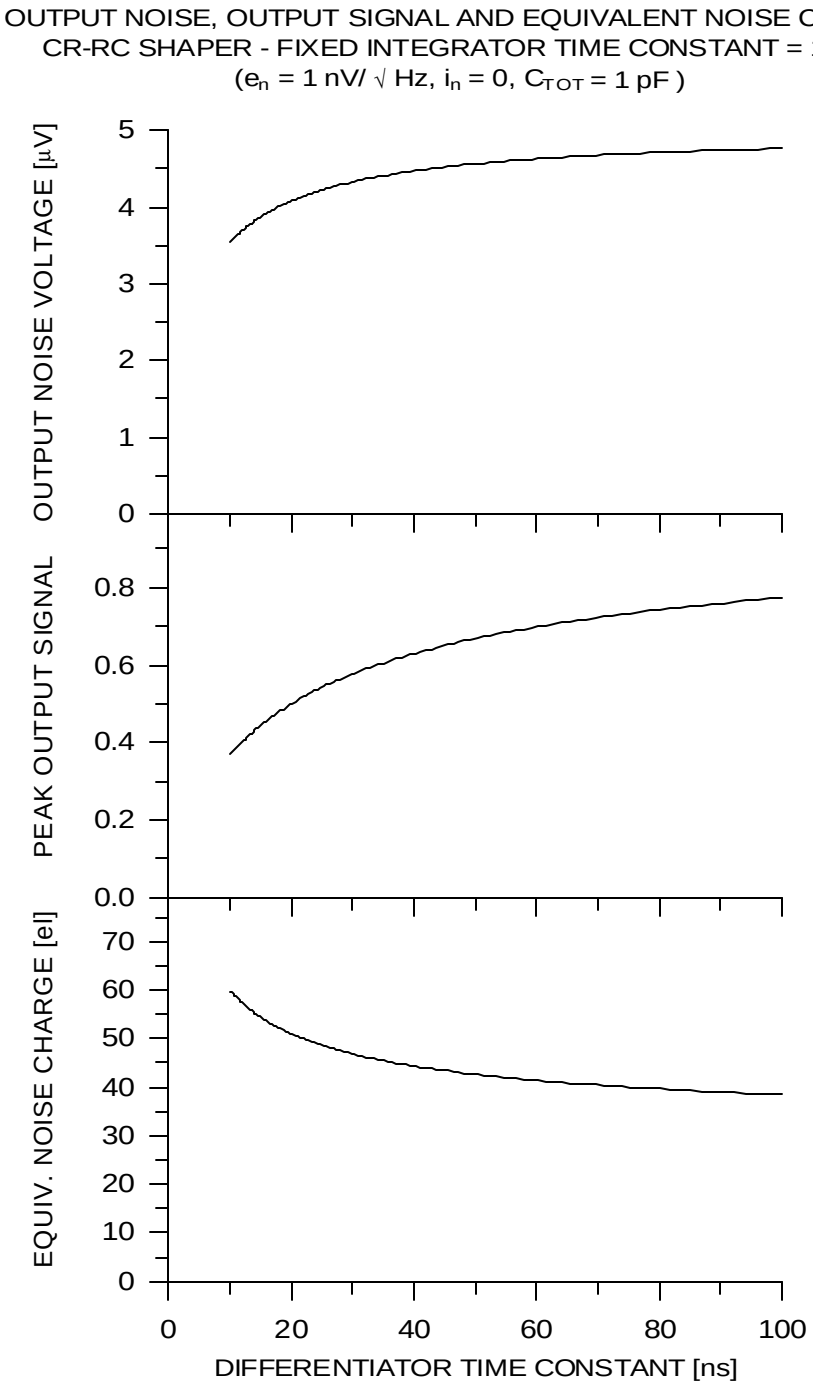
As the differentiator time constant is reduced, the peak signal amplitude at the shaper output decreases.

CR-RC SHAPER
FIXED INTEGRATOR TIME CONSTANT = 10 ns
DIFFERENTIATOR TIME CONSTANT = ∞, 100, 30 and 10 ns



Note that the need to limit the pulse width incurs a significant reduction in the output signal.

Even at a differentiator time constant t_{diff} = 100 ns = 10 t_{int} the output signal is only 80% of the value for $t_{diff} = \infty$, i.e. a system with no low-frequency roll-off.



$$\frac{V_{no}(100\,\text{ns})}{V_{no}(10\,\text{ns})} = 1.3$$

$$\frac{V_{so}(100\,\text{ns})}{V_{so}(10\,\text{ns})} = 2.1$$

Although the noise grows as the differentiator time constant is increased from 10 to 100 ns, it is outweighed by the increase in signal level, so that the net signal-to-noise ratio improves.

$$\frac{Q_n(100\,\text{ns})}{Q_n(10\,\text{ns})} = \frac{1}{1.6}$$

Summary

To evaluate shaper noise performance

- Noise spectrum alone is inadequate

Must also

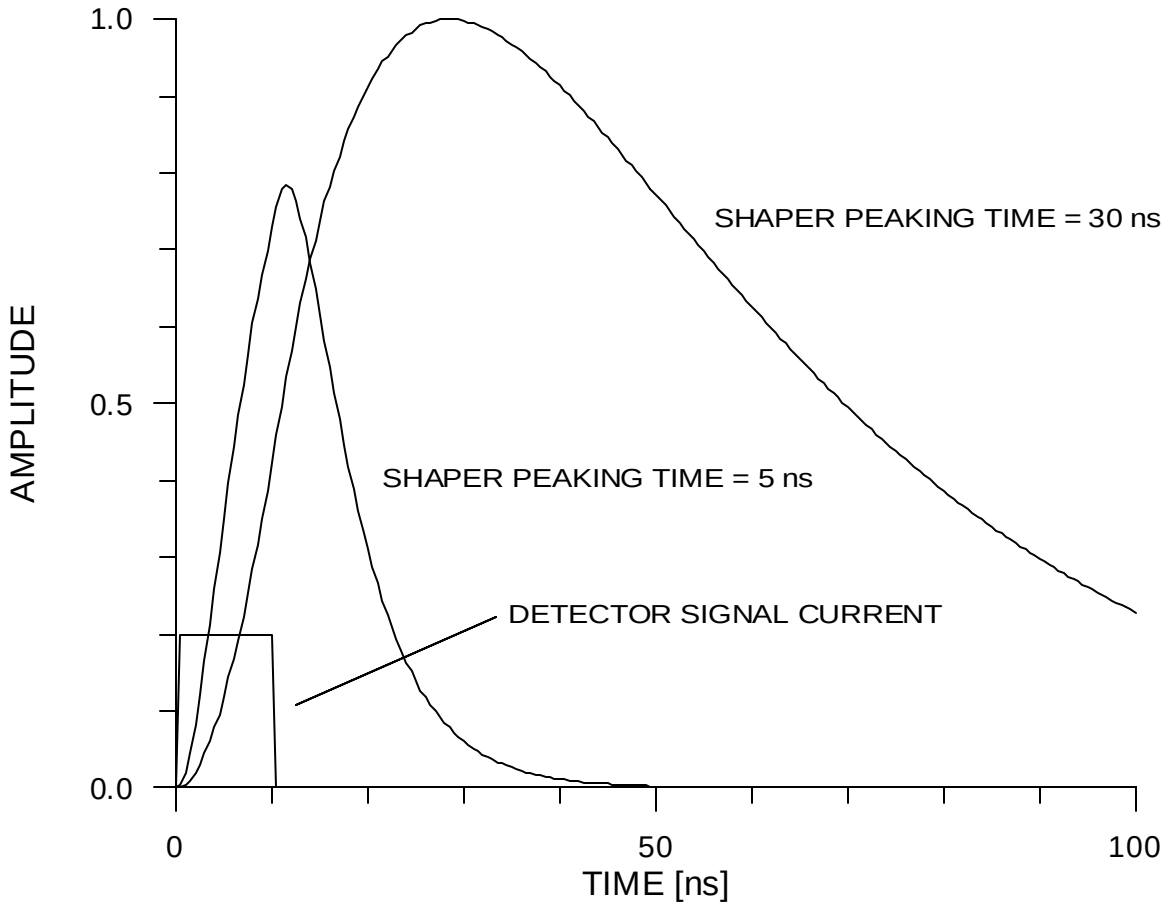
- Assess effect on signal

Signal amplitude is also affected by the relationship of the shaping time to the detector signal duration.

If peaking time of shaper < collection time

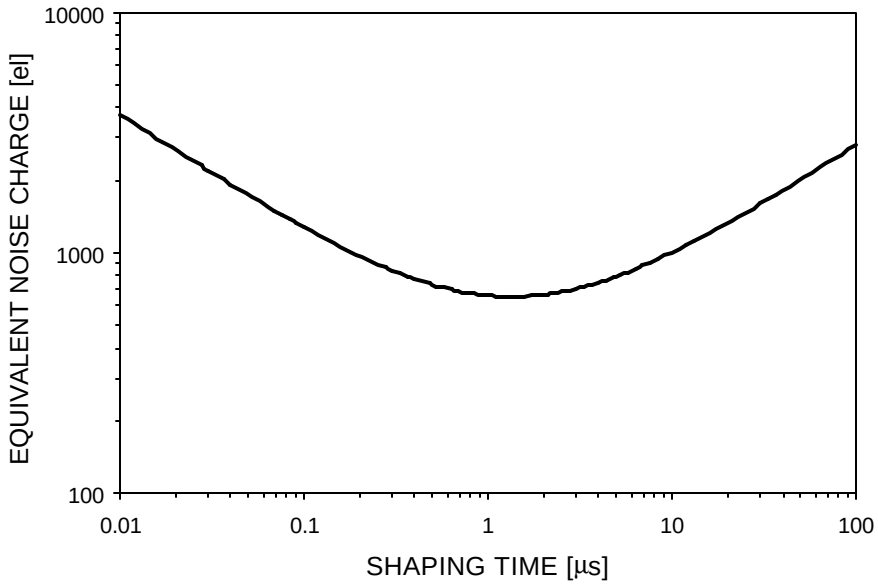
- signal loss (“ballistic deficit”)

Loss in Pulse Height (and Signal-to-Noise Ratio) if Peaking Time of Shaper < Detector Collection Time



Note that although the faster shaper has a peaking time of 5 ns, the response to the detector signal peaks after full charge collection.

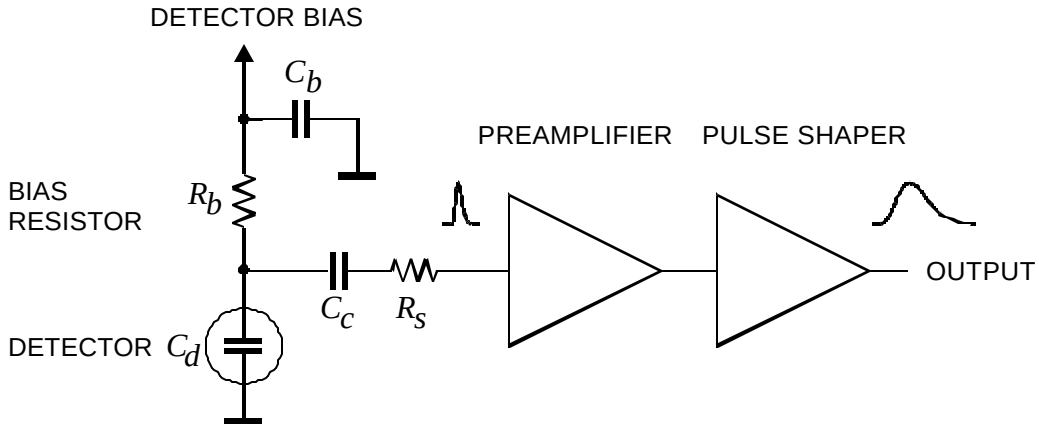
Result of typical noise measurement vs. shaping time



Noise sources (thermal and shot noise) have a flat frequency distribution.

Why doesn't the noise decrease monotonically with increasing shaping time (decreasing bandwidth)?

Analytical Analysis of a Detector Front-End



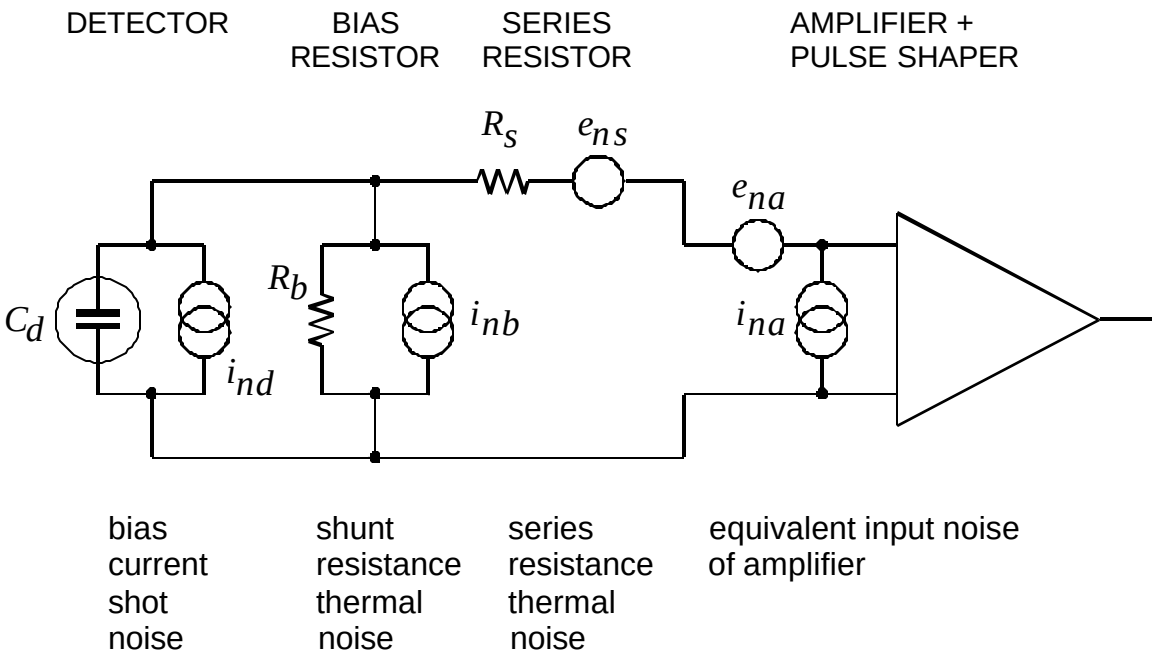
Detector bias voltage is applied through the resistor R_B. The bypass capacitor C_B serves to shunt any external interference coming through the bias supply line to ground. For AC signals this capacitor connects the "far end" of the bias resistor to ground, so that R_B appears to be in parallel with the detector.

The coupling capacitor C_C in the amplifier input path blocks the detector bias voltage from the amplifier input (which is why a capacitor serving this role is also called a "blocking capacitor").

The series resistor R_S represents any resistance present in the connection from the detector to the amplifier input. This includes

- the resistance of the detector electrodes
- the resistance of the connecting wires
- any resistors used to protect the amplifier against large voltage transients ("input protection")
- ... etc.

Equivalent circuit for noise analysis



In this example a voltage-sensitive amplifier is used, so all noise contributions will be calculated in terms of the noise voltage appearing at the amplifier input.

Resistors can be modeled either as voltage or current generators.

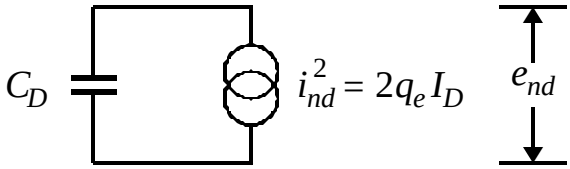
- Resistors in parallel with the input act as current sources
- Resistors in series with the input act as voltage sources.

Steps in the analysis:

1. Determine the frequency distribution of the noise voltage presented to the amplifier input from all individual noise sources
2. Integrate over the frequency response of a CR-RC shaper to determine the total noise output.
3. Determine the output signal for a known signal charge and calculate equivalent noise charge (signal charge for S/N= 1)

Noise Contributions

1. Detector bias current



This model results from two assumptions:

1. The input impedance of the amplifier is infinite

2. The shunt resistance R_P is much larger than the capacitive reactance of the detector in the frequency range of the pulse shaper.

   Does this assumption make sense?

   If R_P is too small, the signal charge on the detector capacitance will discharge before the shaper output peaks. To avoid this

$$R_P C_D >> t_P \approx \frac{1}{w_P}$$

   where w_P is the midband frequency of the shaper. Therefore,

$$R_P >> \frac{1}{w_P C_D}$$

   as postulated.

Under these conditions the noise current will flow through the detector capacitance, yielding the voltage

$$e_{nd}^2 = i_{nd}^2 \frac{1}{(w\,C_D)^2} = 2q_e I_D \frac{1}{(w\,C_D)^2}$$

- **the noise contribution decreases with increasing frequency (shorter shaping time)**

  Note: Although shot noise is "white", the resulting noise spectrum is strongly frequency dependent.

In the time domain this result is more intuitive. Since every shaper also acts as an integrator, one can view the total shot noise as the result of "counting electrons".

Assume an ideal integrator that records all charge uniformly within a time T. The number of electron charges measured is

$$N_e = \frac{I_D T}{q_e}$$

The associated noise is the fluctuation in the number of electron charges recorded

$$s_n = \sqrt{N_e} \propto \sqrt{T}$$

Does this also apply to an AC-coupled system, where no DC current flows, so no electrons are "counted"?

Since shot noise is a fluctuation, the current undergoes both positive and negative excursions. Although the DC component is not passed through an AC coupled system, the excursions are. Since, on the average, each fluctuation requires a positive and a negative zero crossing, the process of "counting electrons" is actually the counting of zero crossings, which in a detailed analysis yields the same result.

2. Parallel Resistance

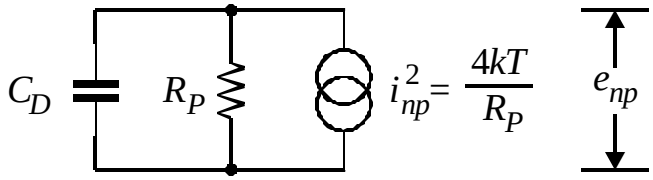
Any shunt resistance R_P acts as a noise current source. In the specific example shown above, the only shunt resistance is the bias resistor R_b.

Additional shunt components in the circuit:

1. bias noise current source
   (infinite resistance by definition)
2. detector capacitance

The noise current flows through both the resistance R_P and the detector capacitance C_D.

- equivalent circuit



The noise voltage applied to the amplifier input is

$$e_{np}^2 = \frac{4kT}{R_P} \left(\frac{R_P \cdot \frac{-\mathbf{i}}{w\ C_D}}{R_P - \frac{\mathbf{i}}{w\ C_D}} \right)^2$$

$$e_{np}^2 = 4kTR_P \frac{1}{1+(w\ R_P C_D)^2}$$

Comment:

Integrating this result over all frequencies yields

$$\int_0^\infty e_{np}^2(w)dw = \int_0^\infty \frac{4kTR_P}{1+(w\,R_PC_D)^2}\,dw = \frac{kT}{C_D}$$

which is independent of R_P. Commonly referred to as "kTC" noise, this contribution is often erroneously interpreted as the "noise of the detector capacitance".

An ideal capacitor has no thermal noise; all noise originates in the resistor.

So, why is the result independent of R_P?

R_P determines the primary noise, but also the noise bandwidth of this subcircuit. As R_P increases, its thermal noise increases, but the noise bandwidth decreases, making the total noise independent of R_P.

However,
If one integrates e_{np} over a bandwidth-limited system

$$E_n^2 = \int_0^\infty 4kTR_P \left|\frac{G(iw)}{1-iw\,R_PC_D}\right|^2 dw$$

the total noise decreases with increasing R_P.

3. Series Resistance

The noise voltage generator associated with the series resistance R_S is in series with the other noise sources, so it simply contributes

$$e_{nr}^2 = 4kTR_S$$

4. Amplifier input noise

The amplifier noise voltage sources usually are not physically present at the amplifier input. Instead the amplifier noise originates within the amplifier, appears at the output, and is referred to the input by dividing the output noise by the amplifier gain, where it appears as a noise voltage generator.

$$e_{na}^2 = e_{nw}^2 + \frac{A_f}{f}$$

$\uparrow$ $\uparrow$

"white noise" $1/f$ noise (can also originate in external components)

This noise voltage generator also adds in series with the other sources.

- Amplifiers generally also exhibit input current noise, which is physically present at the input. Its effect is the same as for the detector bias current, so the analysis given in 1. can be applied.
- In a well-designed amplifier the noise is dominated by the input transistor (fast, high-gain transistors generally best).
  Noise parameters of transistors are discussed in the Appendix.

  Transistor input noise decreases with transconductance
  $\Rightarrow$ increased power
- Minimum device noise limited both by technology and fundamental physics.

Determination of equivalent noise charge

1. Calculate total noise voltage at shaper output
2. Determine peak pulse height at shaper output for a known input charge
3. Input signal for which S/N=1 yields equivalent noise charge

First, assume a simple CR-RC shaper with equal differentiation and integration time constants $t_d = t_i = t$, which in this special case is equal to the peaking time.

The equivalent noise charge

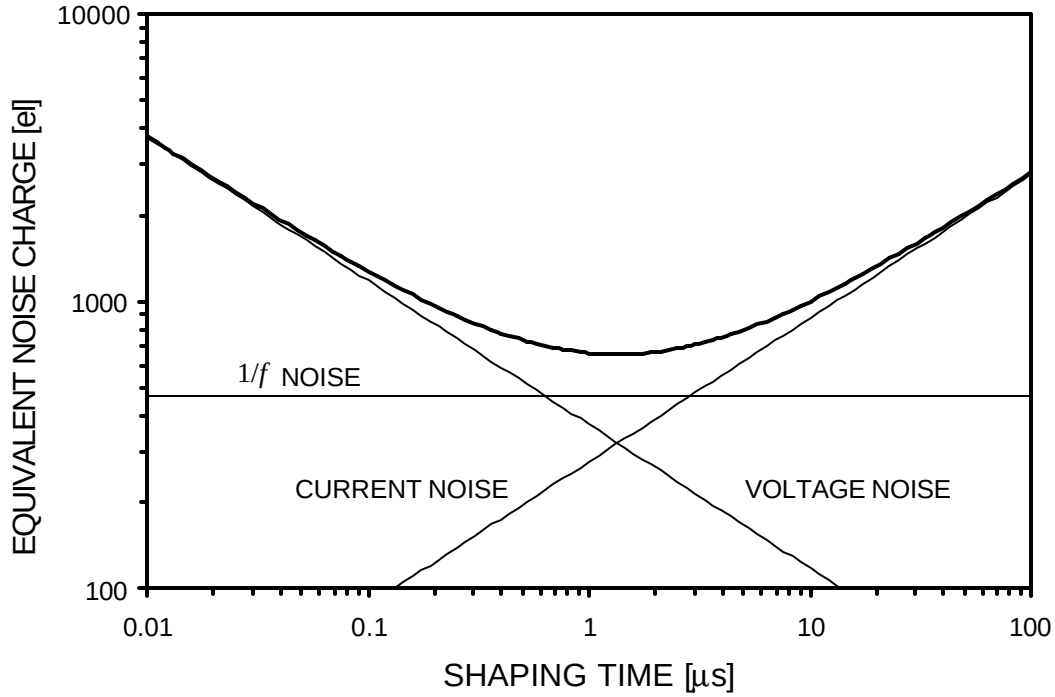
$$Q_n^2 = \left(\frac{e^2}{8}\right)\left[\left(2q_e I_D + \frac{4kT}{R_P} + i_{na}^2\right)\cdot t + \left(4kTR_S + e_{na}^2\right)\cdot\frac{C_D^2}{t} + 4A_f C_D^2\right]$$

current noise	voltage noise	$1/f$ noise
$\propto t$	$\propto 1/t$	independent
independent of C_D	$\propto C_D^2$	of t
		$\propto C_D^2$

- Current noise is independent of detector capacitance, consistent with the notion of “counting electrons”.
- Voltage noise increases with detector capacitance (reduced signal voltage)
- $1/f$ noise is independent of shaping time.
  In general, the total noise of a $1/f$ source depends on the ratio of the upper to lower cutoff frequencies, not on the absolute noise bandwidth. If t_d and t_i are scaled by the same factor, this ratio remains constant.
- detector leakage current and FET noise decrease with temperature
  $\Rightarrow$ high resolution Si and Ge detectors for x-rays and gamma rays operate at cryogenic temperatures.

The equivalent noise charge Q_n assumes a minimum when the current and voltage noise contributions are equal.

Typical Result



dominated by voltage noise current noise

For a CR-RC shaper the noise minimum obtains for $t_d = t_i = t$.

This criterion does not hold for more sophisticated shapers.

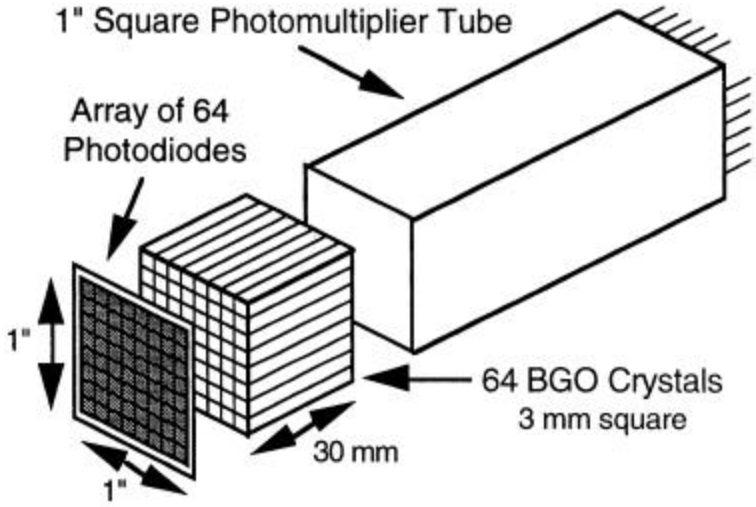
Caution: Even for a CR-RC shaper this criterion only applies when the differentiation time constant is the primary parameter, i.e. when the pulse width must be constrained.

When the rise time, i.e. the integration time constant, is the primary consideration, it is advantageous to make $t_d > t_i$, since the signal will increase more rapidly than the noise, as was shown previously

Example: Photodiode Readout

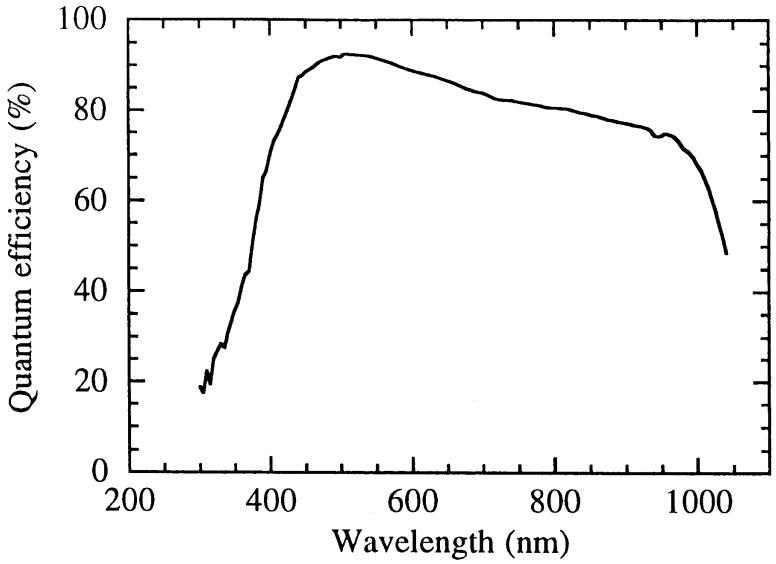
(S. Holland, N. Wang, I. Kipnis, B. Krieger, W. Moses, LBNL)

Medical Imaging (Positron Emission Tomography)



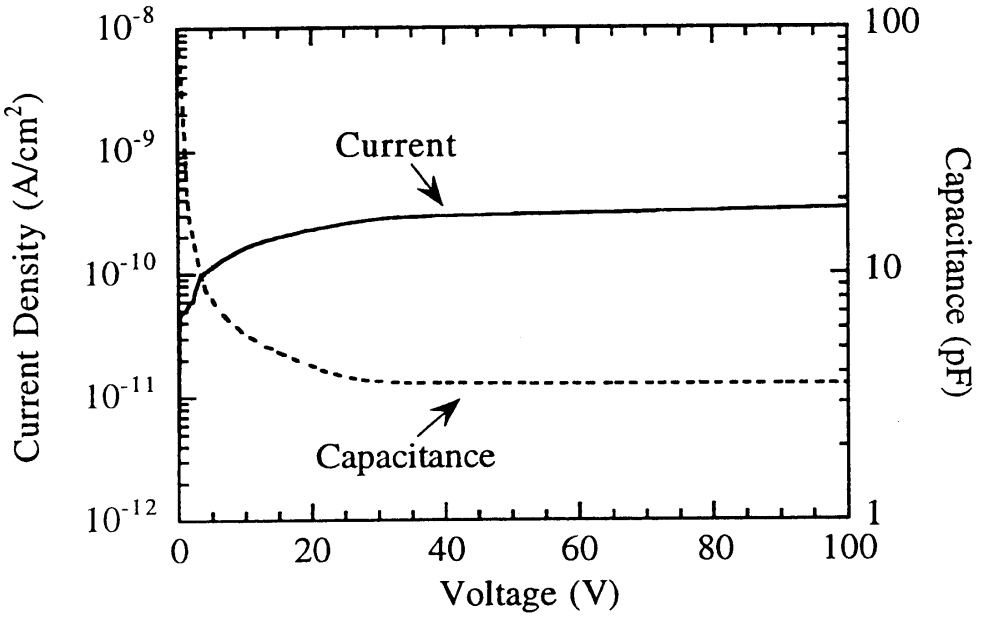
Read out 64 BGO crystals with one PMT (timing, energy) and tag crystal by segmented photodiode array.

Requires thin dead layer on photodiode to maximize quantum efficiency.



Thin electrode must be implemented with low resistance to avoid significant degradation of electronic noise.

Furthermore, low reverse bias current critical to reduce noise.



Photodiodes designed and fabricated in LBNL Microsystems Lab.

Front-end chip (preamplifier + shaper):

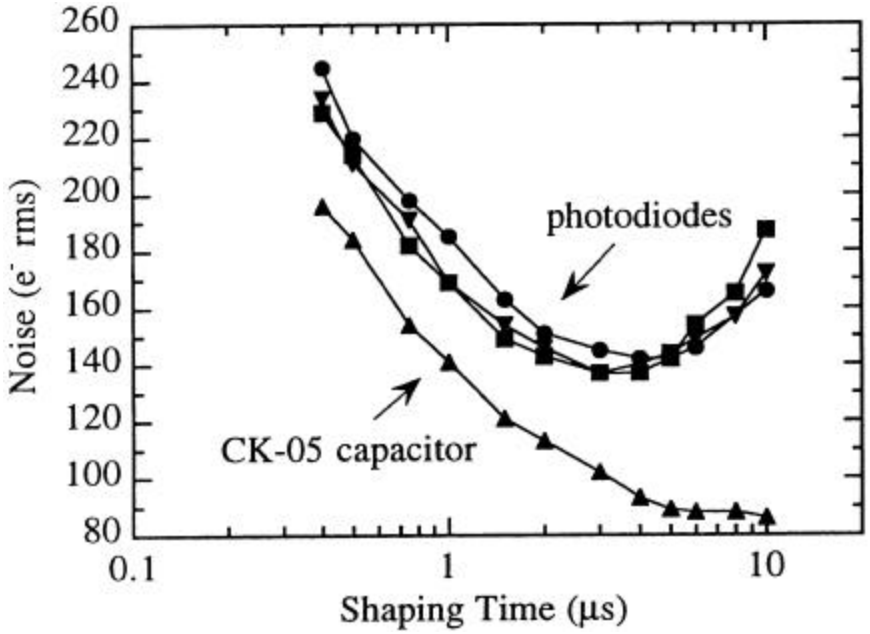
16 channels per chip

die size: 2 x 2 mm^2,
1.2 µm CMOS

continuously adjustable shaping time (0.5 to 50 µs)

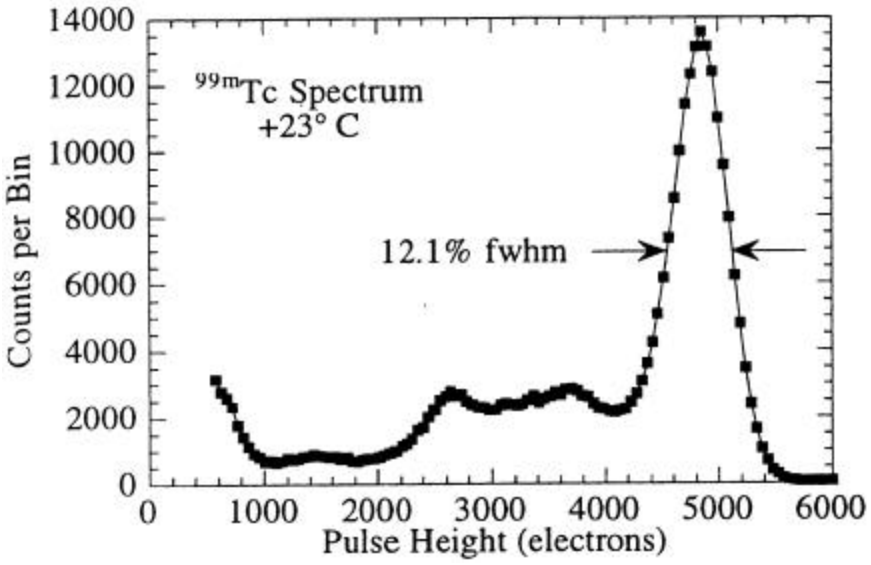
gain: 100 mV per 1000 el.

Noise vs. shaping time



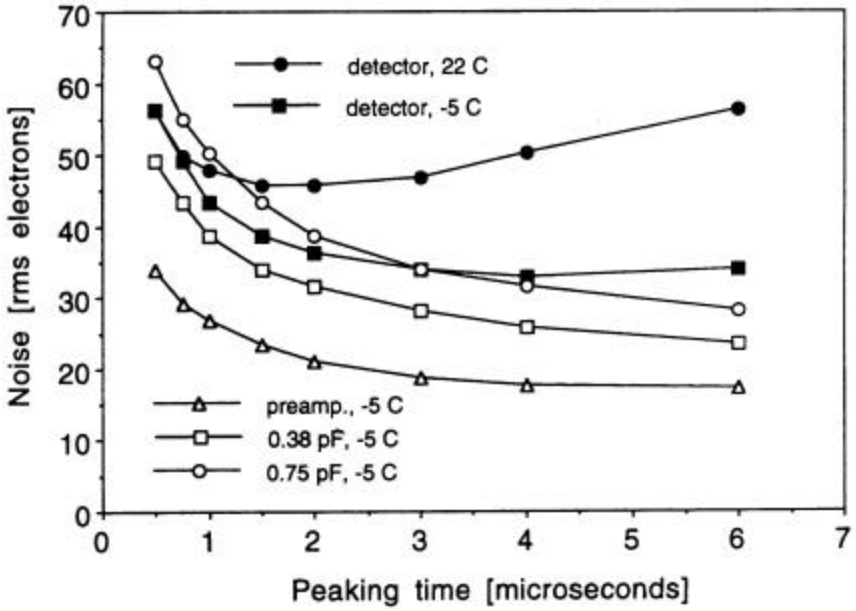
Note increase in noise at long shaping times when photodiode is connected - shot noise contribution.

Energy spectrum with BGO scintillator



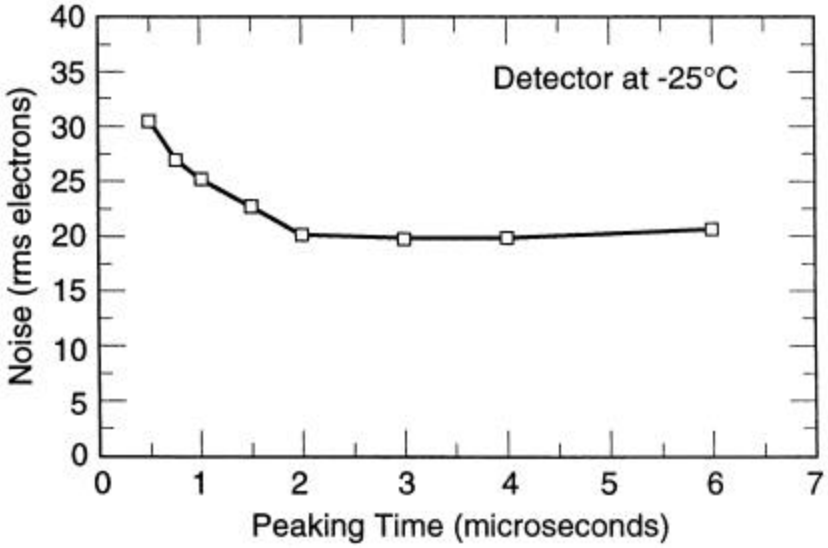
Another Example: short-strip Si x-ray detector

(B. Ludewigt, C. Rossington, I. Kipnis, B. Krieger, LBNL)



- Connecting the detector increases noise because of added capacitance and detector current (as indicated by increase of noise with peaking time).
- Cooling the detector reduces the current and noise improves.

Second prototype



Current noise negligible because of cooling –
"flat" noise vs. shaping time indicates that 1/*f* noise dominates.

Numerical expression for the noise of a CR-RC shaper
(amplifier current noise negligible)

(note that some units are "hidden" in the numerical factors)

$$Q_n^2 = 12\, I_B\, t \;+\; 6\cdot 10^5 \frac{t}{R_P} + \; 3.6\cdot 10^4\, e_n^2 \frac{C^2}{t} \quad [\text{rms electrons}^2]$$

where

t shaping time constant [ns]

I_B detector bias current + amplifier input current [nA]

R_P input shunt resistance [kΩ]

e_n equivalent input noise voltage spectral density [nV/√Hz]

C total input capacitance [pF]

Q_n= 1 *el* corresponds to 3.6 eV in Si
2.9 eV in Ge

Note:

For sources connected in parallel, currents are additive.

For sources connected in series, voltages are additive.

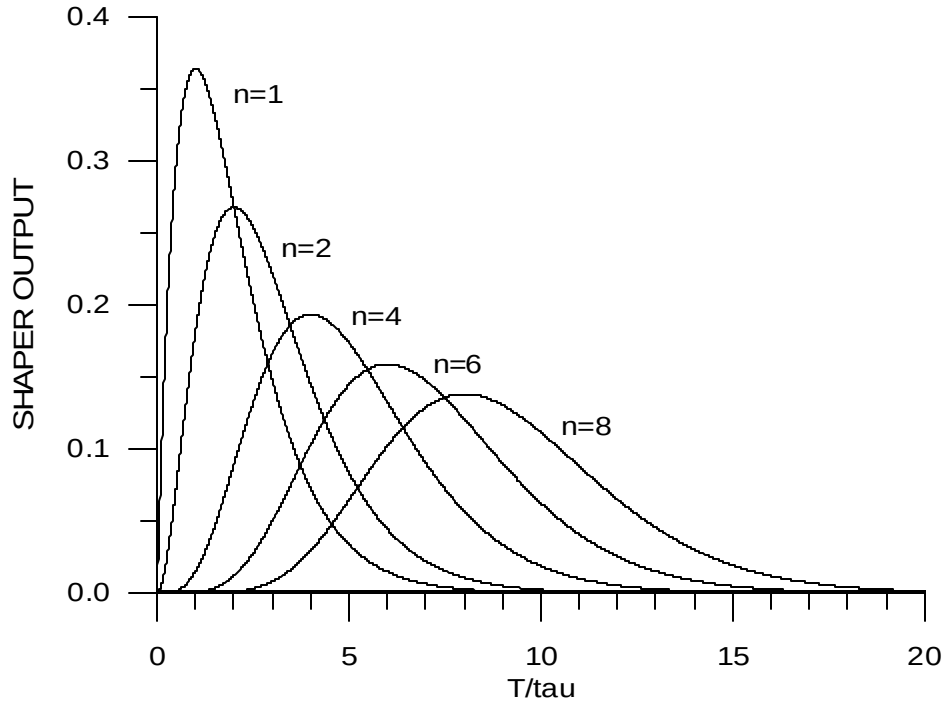
- In the detector community voltage and current noise are often called “series” and “parallel” noise.

  The rest of the world uses equivalent noise voltage and current.

  Since they are physically meaningful, use of these widely understood terms is preferable.

CR-RC Shapers with Multiple Integrators

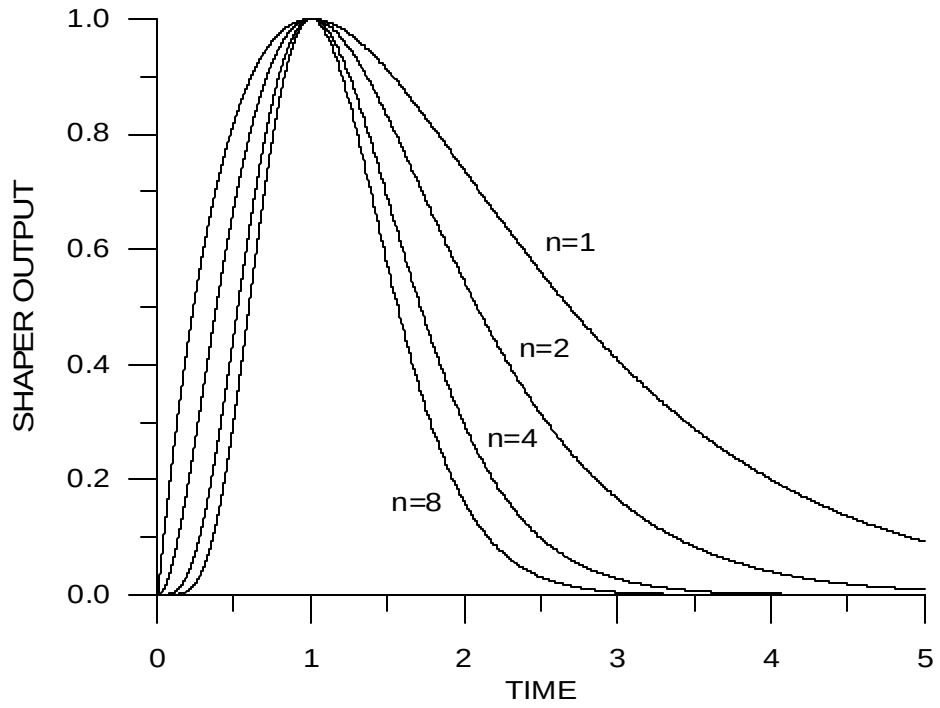
a. Start with simple CR-RC shaper and add additional integrators (n= 1 to n= 2, ... n= 8) with the same time constant t .



With additional integrators the peaking time T_p increases

$$T_p = n\,t$$

b) Time constants changed to preserve the peaking time
($t_n = t_{n=1}/n$)



Increasing the number of integrators makes the output pulse more symmetrical with a faster return to baseline.

- improved rate capability at the same peaking time

Shapers with the equivalent of 8 RC integrators are common. Usually, this is achieved with active filters (i.e. circuitry that synthesizes the bandpass with amplifiers and feedback networks).

Noise Analysis in the Time Domain

The noise analysis of shapers is rather straightforward if the frequency response is known.

On the other hand, since we are primarily interested in the pulse response, shapers are often designed directly in the time domain, so it seems more appropriate to analyze the noise performance in the time domain also.

Clearly, one can take the time response and Fourier transform it to the frequency domain, but this approach becomes problematic for time-variant shapers.

The CR-RC shapers discussed up to now utilize filters whose time constants remain constant during the duration of the pulse, i.e. they are time-invariant.

Many popular types of shapers utilize signal sampling or change the filter constants during the pulse to improve pulse characteristics, i.e. faster return to baseline or greater insensitivity to variations in detector pulse shape.

These time-variant shapers cannot be analyzed in the manner described above. Various techniques are available, but some shapers can be analyzed only in the time domain.

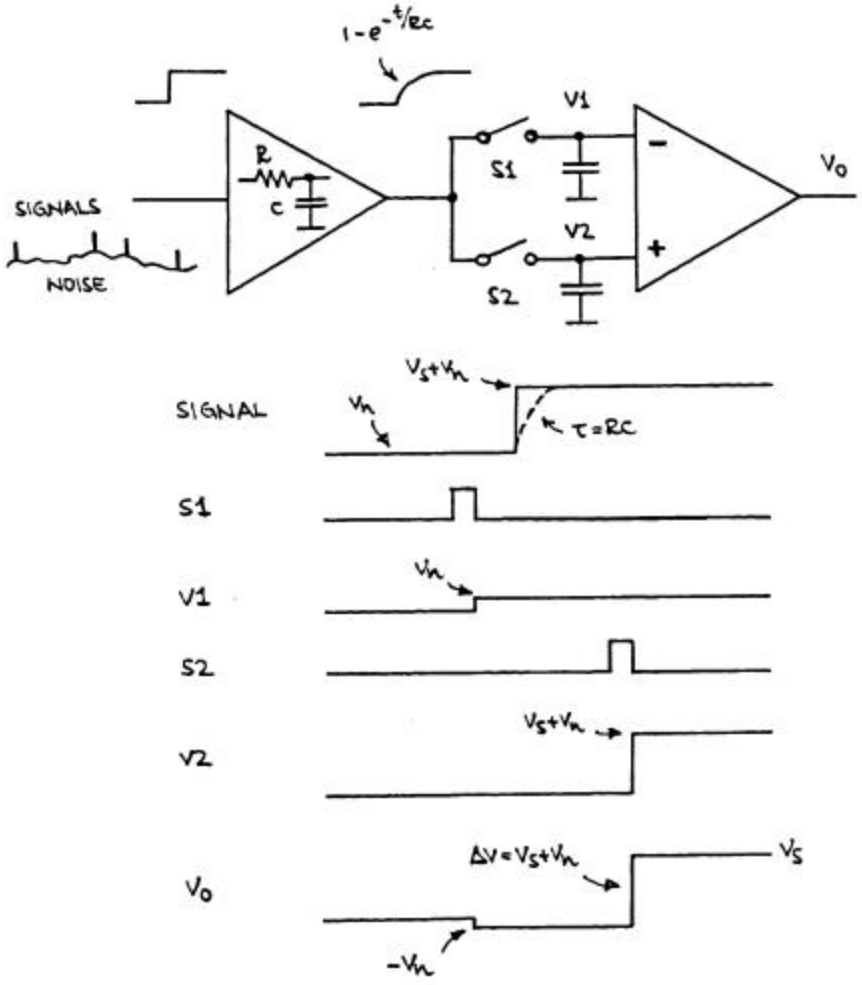
References:

V. Radeka, Nucl. Instr. and Meth. **99** (1972) 525
V. Radeka, IEEE Trans. Nucl. Sci. **NS-21** (1974) 51
F.S. Goulding, Nucl. Instr. and Meth. **100** (1972) 493
F.S. Goulding, IEEE Trans. Nucl. Sci. **NS-29** (1982) 1125

Example:

A commonly used time-variant filter is the correlated double-sampler.

This shaper can be analyzed exactly only in the time domain.



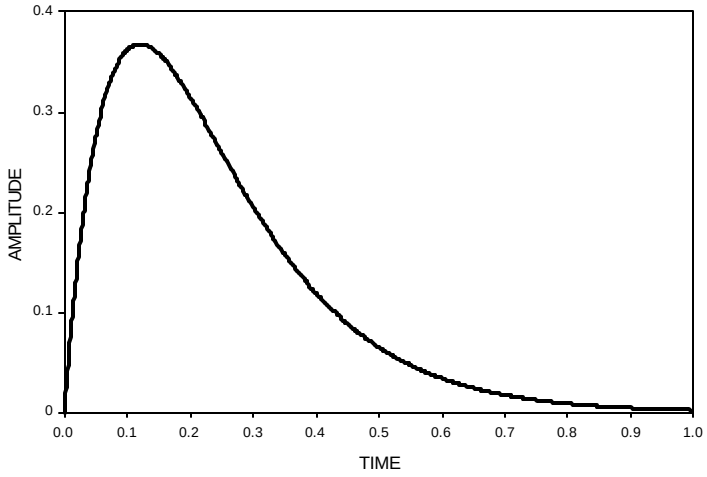
1. Signals are superimposed on a (slowly) fluctuating baseline
2. To remove baseline fluctuations the baseline is sampled prior to the arrival of a signal.
3. Next, the signal + baseline is sampled and the previous baseline sample subtracted to obtain the signal

S/N depends on
1. time constant of prefilter
2. time difference between samples

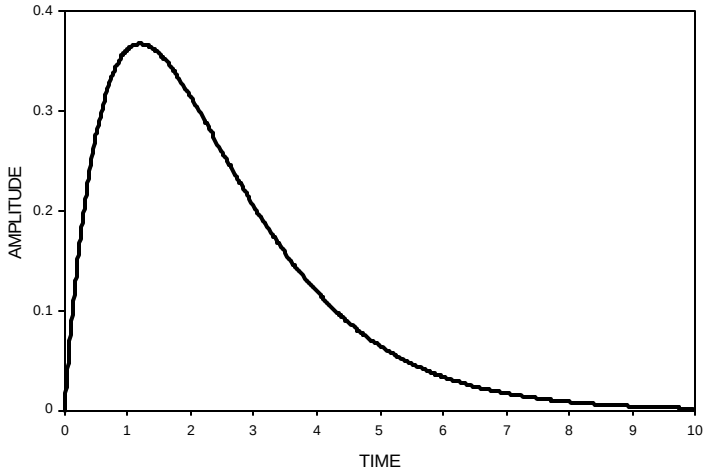
Scaling of Filter Parameters

Pulse shape is the same when shaping time is changed

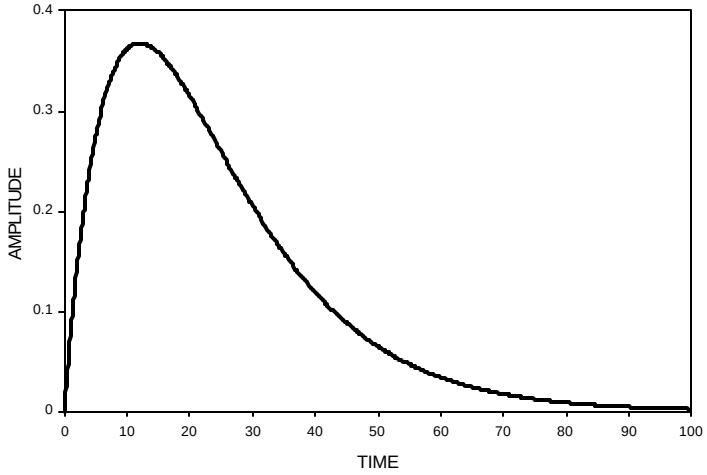
shaping time = t



shaping time = 10 t



shaping time = 100 t



Since the pulse width is directly related to the noise bandwidth (Parseval's Theorem),

$$\int_0^\infty |A(f)|^2 \, df = \int_{-\infty}^{\infty} [F(t)]^2 \, dt \ ,$$

the noise charge

$$Q_n^2 = \left(\frac{e^2}{8}\right)\left[\left(2q_e I_D + \frac{4kT}{R_P} + i_{na}^2\right)\cdot t + \left(4kTR_S + e_{na}^2\right)\cdot\frac{C_D^2}{t} + 4A_f C_D^2\right]$$

can be written in a general form that applies to any shaper.

$$Q_n^2 = \ i_n^2 \ T \ F_i \ + \ C^2 \ e_n^2 \ \frac{1}{T} \ F_v \ + \ C^2 \ A_f \ F_{vf}$$

The individual current and voltage noise contributions are combined:

current noise $\quad i_n^2 = 2q_e I_b + \frac{4kT}{R_P} + i_{na}^2$

and voltage noise $\quad e_n^2 = 4kTR_S + e_{na}^2$

The shaper is characterized by noise coefficients F_i, F_v and F_{vf}, which depend only on the shape of the pulse.

The noise bandwidth scales with a characteristic time T.

In the specific case of a CR-RC shaper T is equal to the peaking time T_p, the time at which the pulse assumes its maximum value. For a correlated double sampler, the sampling time is an appropriate measure.

The first term describes the current noise contribution, whereas the second and third terms describe the voltage noise contributions due to white and $1/f$ noise sources.

- Generally, the noise indices or "shape factors" F_i, F_v and F_{vf} characterize the type of shaper, for example CR-RC or CR-$(RC)^4$.
- They depend only on the ratio of time constants t_d/t_i, rather than their absolute magnitude.
- The noise contribution then scales with the characteristic time T. The choice of characteristic time is somewhat arbitrary. so any convenient measure for a given shaper can be adopted in deriving the noise coefficients F.

The shape factors F_i, F_v are easily calculated

$$F_i = \frac{1}{2T_S}\int_{-\infty}^{\infty}\left[W(t)\right]^2 dt\ , \qquad F_v = \frac{T_S}{2}\int_{-\infty}^{\infty}\left[\frac{dW(t)}{dt}\right]^2 dt$$

For time invariant pulse shaping $W(t)$ is simply the system's impulse response, with the peak output signal normalized to unity.

Recipe:

Inject a short current pulse injected into the preamplifier

For a charge-sensitive preamp this is generated by a voltage step applied to the test input.

$W(t)$ is the output signal as seen on an oscilloscope. With a digitizing oscilloscope the signal can be recorded and numerically normalized, squared, and integrated.

Summary

Two basic noise mechanisms: input noise current i_n
input noise voltage e_n

Equivalent Noise Charge:

$$Q_n^2 = i_n^2 \, T_s \, F_i + C^2 \, e_n^2 \, \frac{F_v}{T_s}$$

where T_s Characteristic shaping time (*e.g.* peaking time)

F_i , F_v "Shape Factors" that are determined by the shape of the pulse.

C Total capacitance at the input node (detector capacitance + input capacitance of preamplifier + stray capacitance + …)

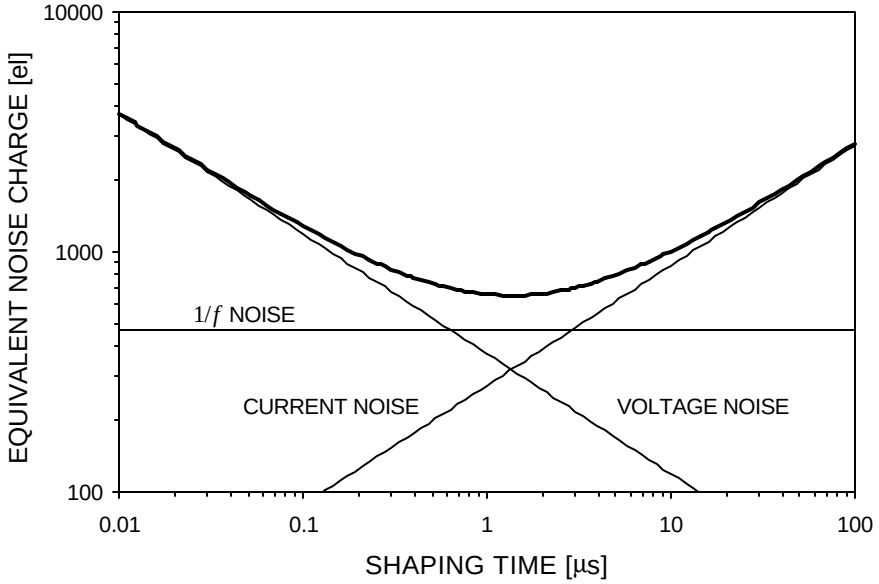
Typical values of F_i , F_v

CR-RC shaper	F_i = 0.924	F_v = 0.924
CR-(RC)^4 shaper	F_i = 0.45	F_v = 1.02
CR-(RC)^7 shaper	F_i = 0.34	F_v = 1.27
CAFE chip	F_i = 0.4	F_v = 1.2

Note that $F_i < F_v$ for higher order shapers.

Shapers can be optimized to reduce current noise contribution relative to the voltage noise (mitigate radiation damage!).

Minimum noise obtains when the current and voltage noise contributions are equal.



Current noise

- detector bias current
  increases with detector size, strongly temperature dependent
- resistors shunting the input
  increases as resistance is decreased
- input transistor – low for FET, higher for BJTs

Voltage noise

- input transistor (see Appendix)
- series resistance
  e.g. detector electrode, protection circuits

FETs commonly used as input devices
improved noise performance when cooled ($T_{opt} \approx 130$ K)

Bipolar transistors advantageous at short shaping times (<100 ns).
When collector current is optimized, bipolar transistor equivalent noise charge is independent of shaping time (see Appendix).

Equivalent Noise Charge vs. Detector Capacitance ($C = C_d + C_a$)

$$Q_n = \sqrt{i_n^2 F_i T + (C_d + C_a)^2 e_n^2 F_v \frac{1}{T}}$$

$$\frac{dQ_n}{dC_d} = \frac{2C_d e_n^2 F_v \frac{1}{T}}{\sqrt{i_n^2 F_i T + (C_d + C_a)^2 e_n^2 F_v \frac{1}{T}}}$$

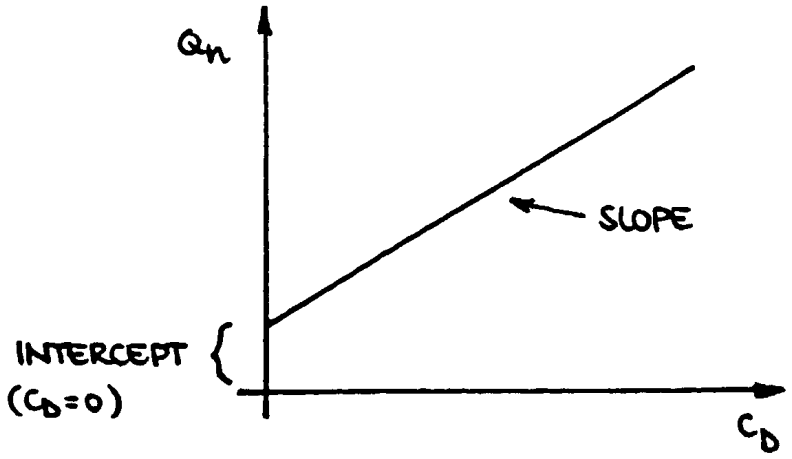
If current noise $i_n^2 F_i T$ is negligible, i.e. **voltage noise dominates**

$$\frac{dQ_n}{dC_d} \approx 2e_n \cdot \sqrt{\frac{F_v}{T}}$$

↑ input stage ↑ shaper

Zero intercept

$$Q_n\big|_{C_d=0} = C_a e_n \sqrt{F_v / T}$$



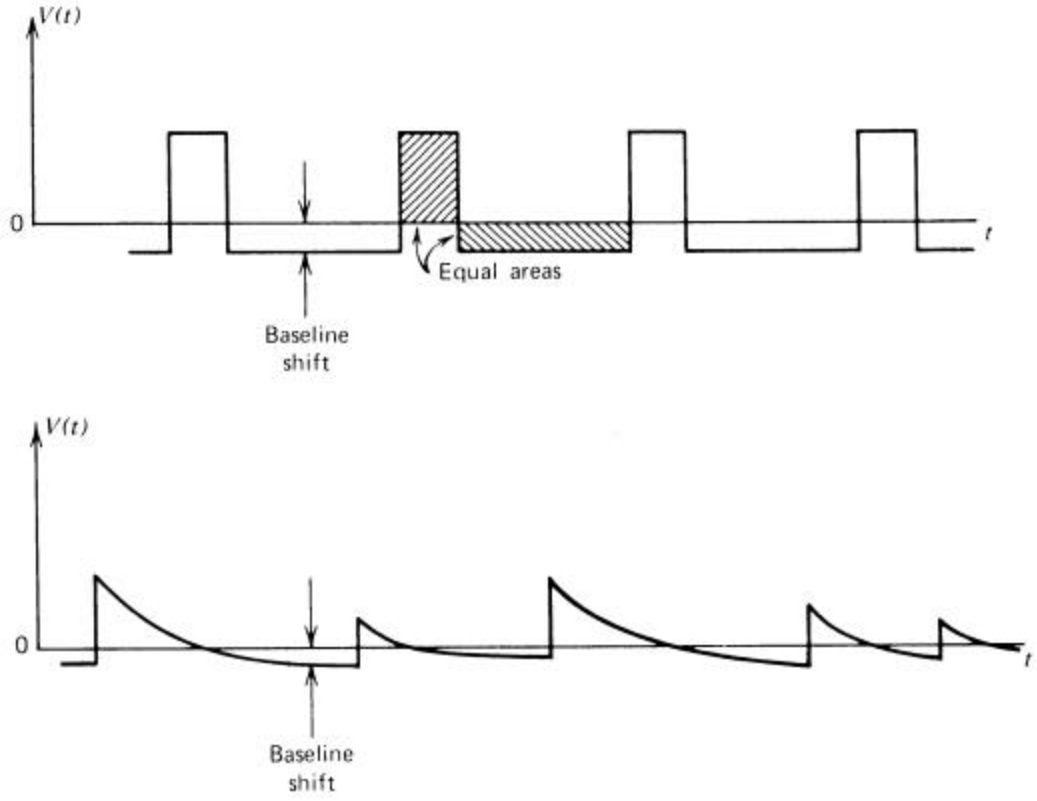
4. Some Other Aspects of Pulse Shaping

4.1 Baseline Restoration

Any series capacitor in a system prevents transmission of a DC component.

A sequence of unipolar pulses has a DC component that depends on the duty factor, i.e. the event rate.

- The baseline shifts to make the overall transmitted charge equal zero.



(from Knoll)

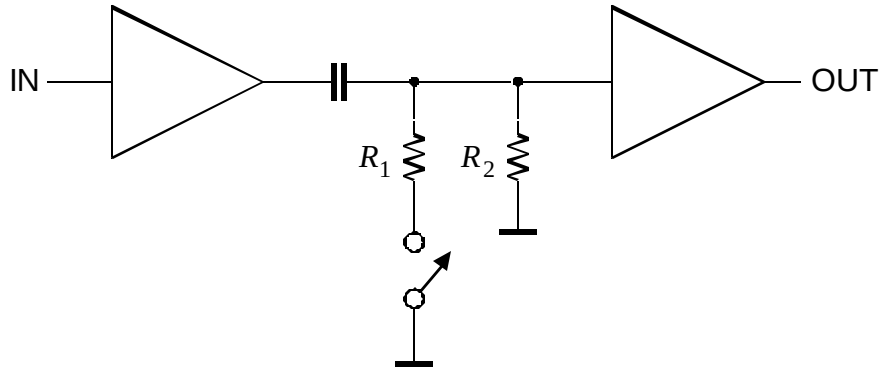
Random rates lead to random fluctuations of the baseline shift

- spectral broadening
- These shifts occur whenever the DC gain is not equal to the midband gain

The baseline shift can be mitigated by a baseline restorer (BLR).

Principle of a baseline restorer:

Connect signal line to ground during the absence of a signal to establish the baseline just prior to the arrival of a pulse.



R_1 and R_2 determine the charge and discharge time constants. The discharge time constant (switch opened) must be much larger than the pulse width.

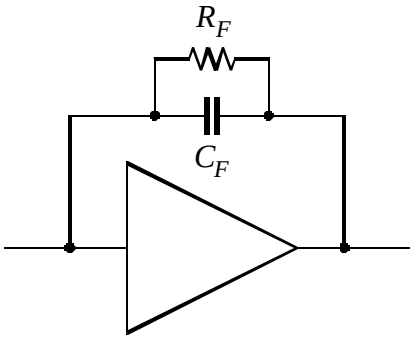
Originally performed with diodes (passive restorer), baseline restoration circuits now tend to include active loops with adjustable thresholds to sense the presence of a signal (gated restorer). Asymmetric charge and discharge time constants improve performance at high count rates.

- This is a form of time-variant filtering. Care must be exercised to reduce noise and switching artifacts introduced by the BLR.

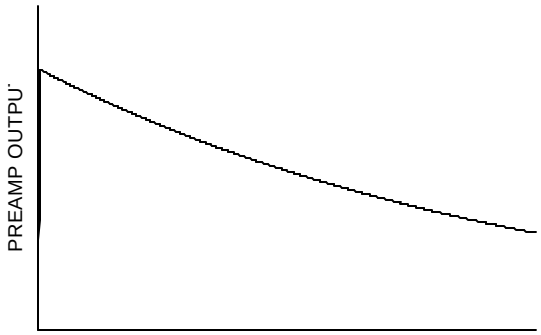
- Good pole-zero cancellation (next topic) is crucial for proper baseline restoration.

4.2 Pole Zero Cancellation

Feedback capacitor in charge sensitive preamplifier must be discharged. Commonly done with resistor.



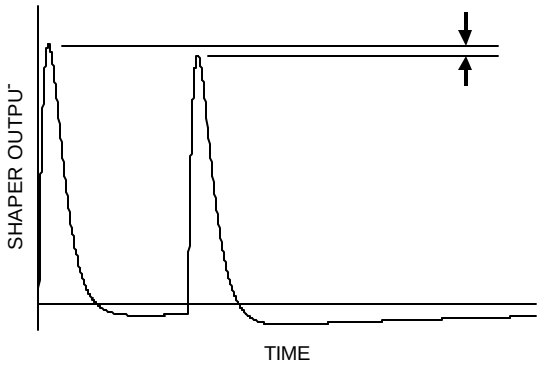
Output no longer a step, but decays exponentially



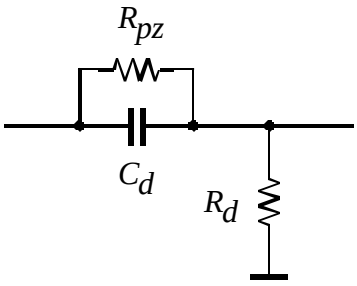
Exponential decay superimposed on shaper output.

$\Rightarrow$ undershoot

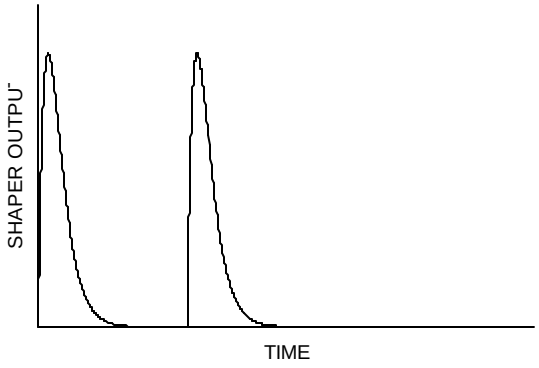
$\Rightarrow$ loss of resolution due to baseline variations



Add R_{pz} to differentiator:



"zero" cancels "pole" of preamp when $R_F C_F = R_{pz} C_d$



Not needed in pulsed reset circuits (optical or transistor)

Technique also used to compensate for "tails" of detector pulses: "tail cancellation"

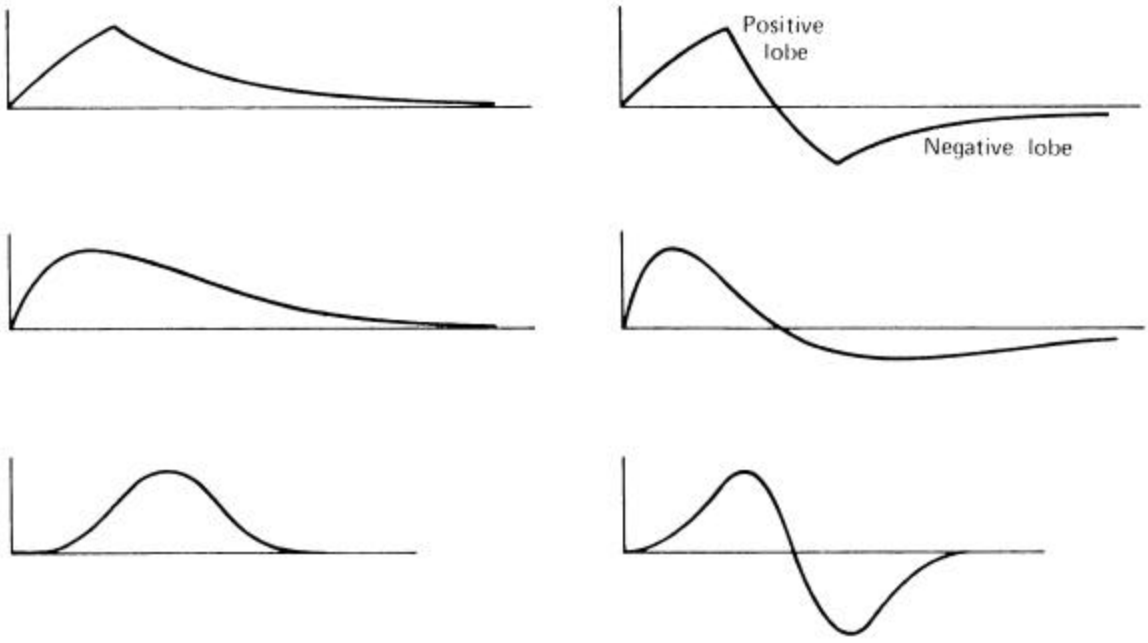
Critical for proper functioning of baseline restorer.

4.3 Bipolar vs. Unipolar Shaping

Unipolar pulse + 2nd differentiator ® Bipolar pulse

Examples:

unipolar bipolar



Electronic resolution with bipolar shaping typ. 25 – 50% worse than for corresponding unipolar shaper.

However …

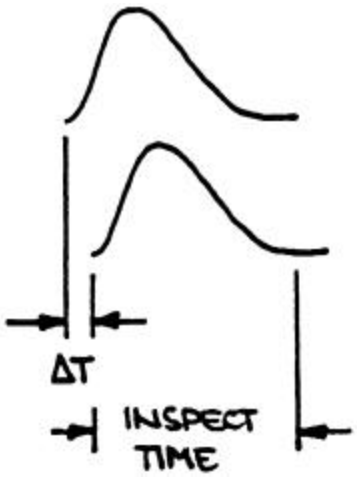
- Bipolar shaping eliminates baseline shift (as the DC component is zero).
- Pole-zero adjustment less critical
- Added suppression of low-frequency noise (see Part 7).
- Not all measurements require optimum noise performance. Bipolar shaping is much more convenient for the user (important in large systems!) – often the method of choice.

4.4 Pulse Pile-Up and Pile-Up Rejectors

pile-up ⇒ false amplitude measurement

Two cases:

1.



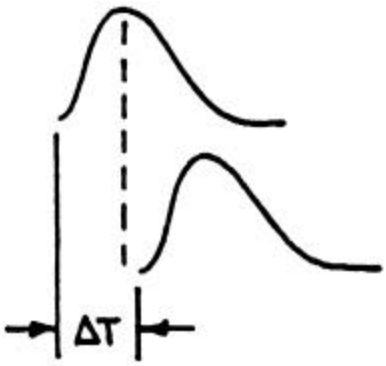
ΔT < time to peak

Both peak amplitudes are affected by superposition.

⇒ Reject both pulses

Dead Time: ΔT + inspect time
(~ pulse width)

2.



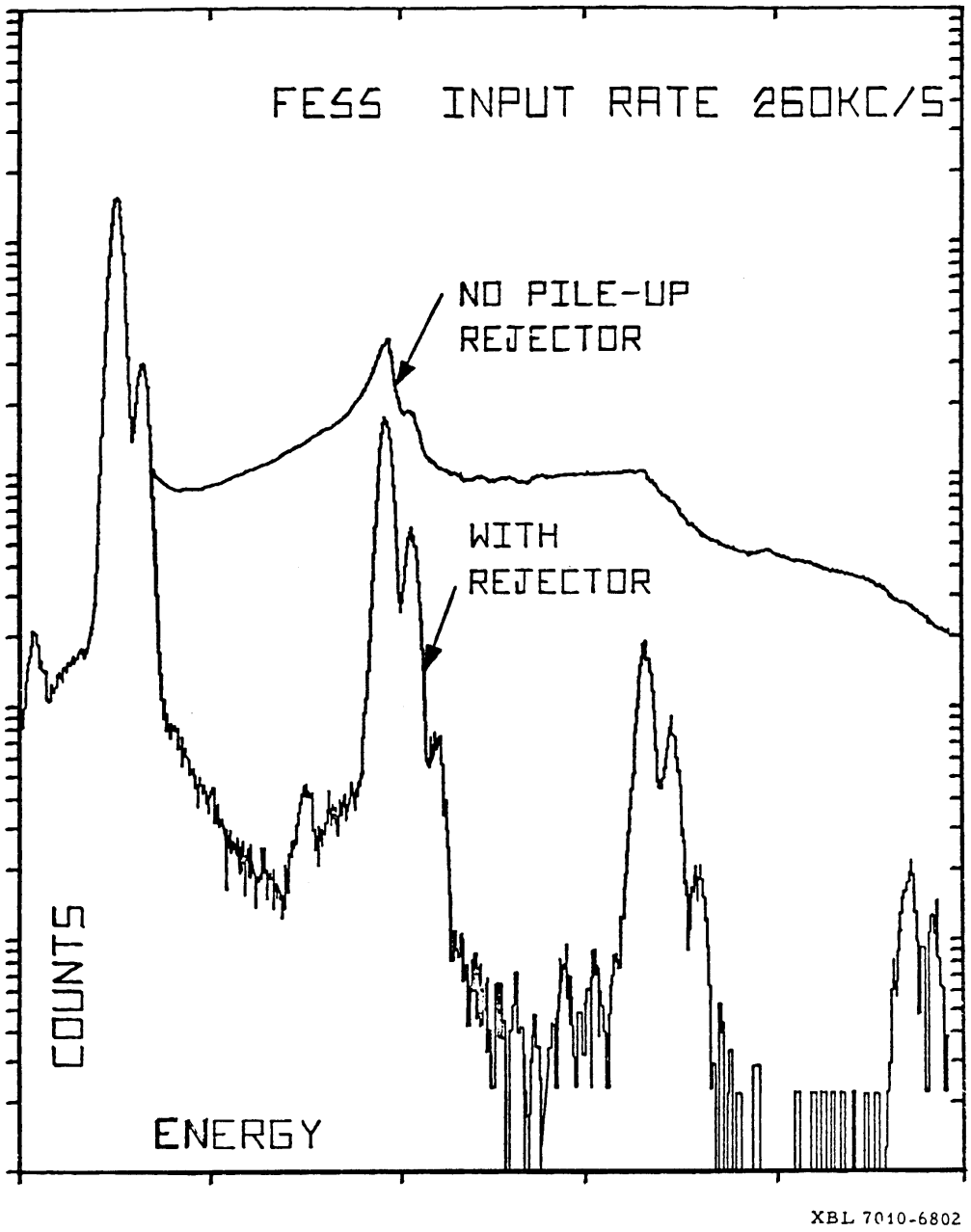
ΔT > time to peak and
ΔT < inspect time, i.e.
time where amplitude of first pulse << resolution

Peak amplitude of first pulse unaffected.

⇒ Reject 2nd pulse only

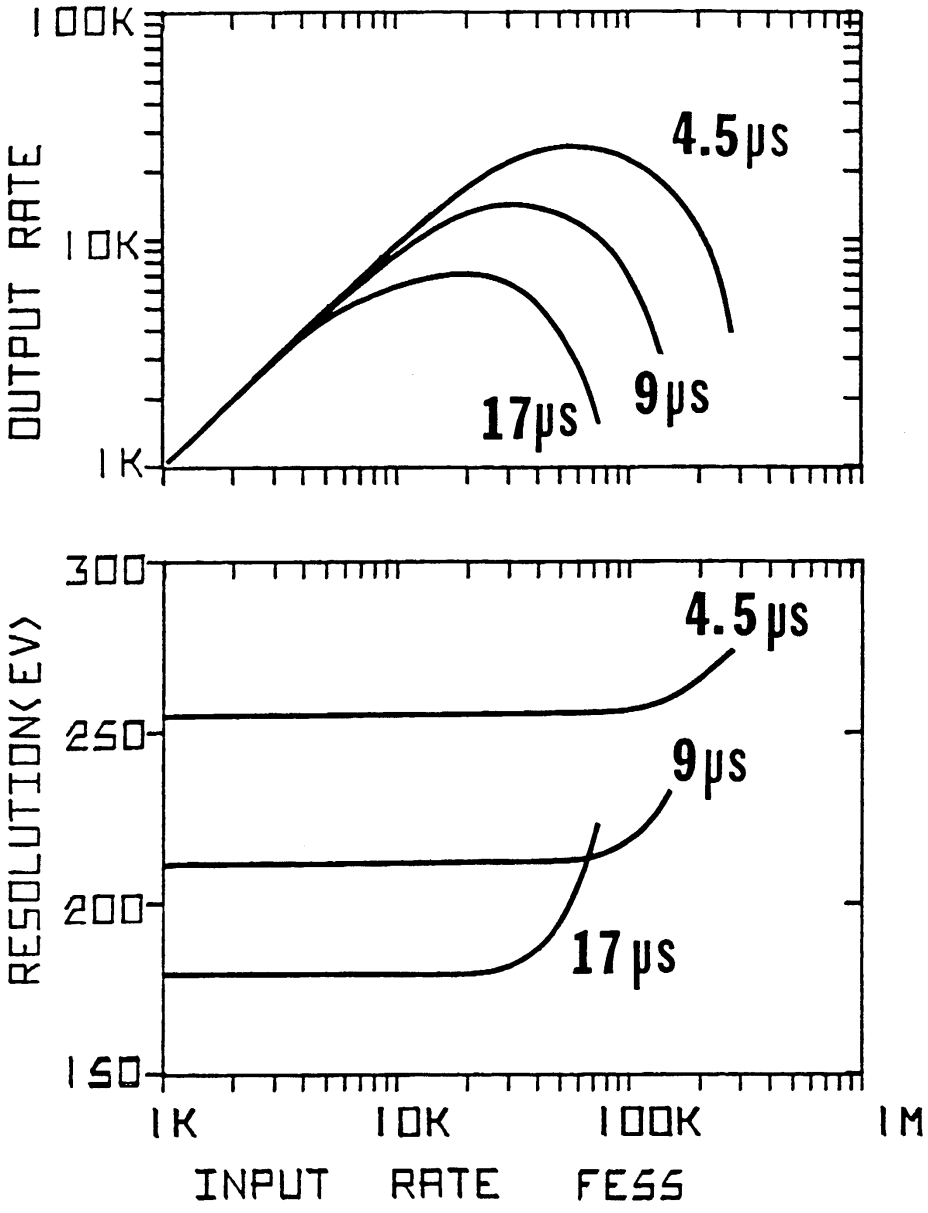
No additional dead time if first pulse accepted for digitization and dead time of ADC >
(DT + inspect time)

Typical Performance of a Pile-Up Rejector



(Don Landis)

Dead Time and Resolution vs. Counting Rate



(Joe Jaklevic)

Throughput peaks and then drops as the input rate increases, as most events suffer pile-up and are rejected.

Resolution also degrades beyond turnover point.

- Turnover rate depends on pulse shape and PUR circuitry.
- Critical to measure throughput vs. rate!

Limitations of Pile-Up Rejectors

Minimum dead time where circuitry can't recognize second pulse

⇒ spurious sum peaks

Detectable dead time depends on relative pulse amplitudes

e.g. small pulse following large pulse

⇒ amplitude-dependent rejection factor

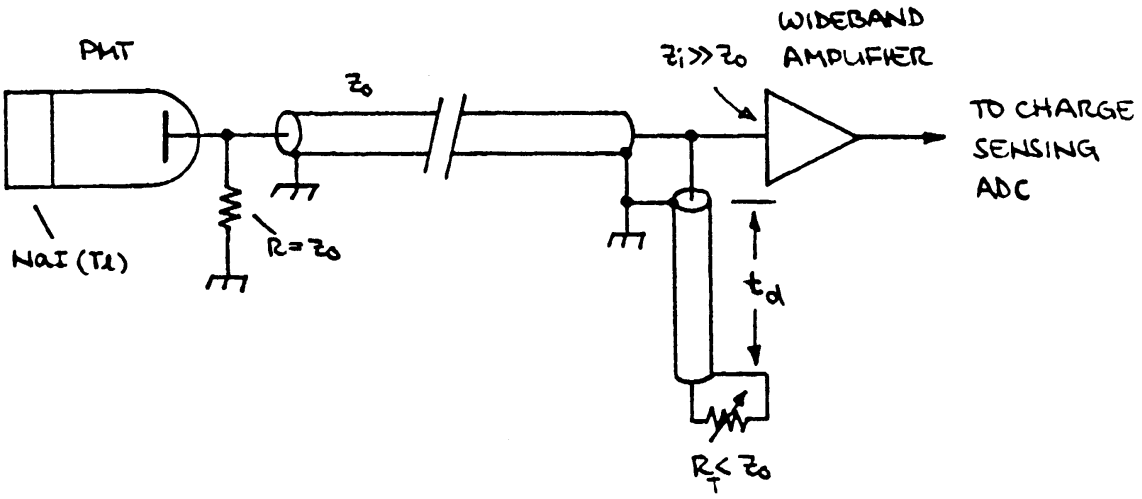
problem when measuring yields!

These effects can be evaluated and taken into account, but in experiments it is often appropriate to avoid these problems by using a shorter shaping time (trade off resolution for simpler analysis).

4.5 Delay-Line Clipping

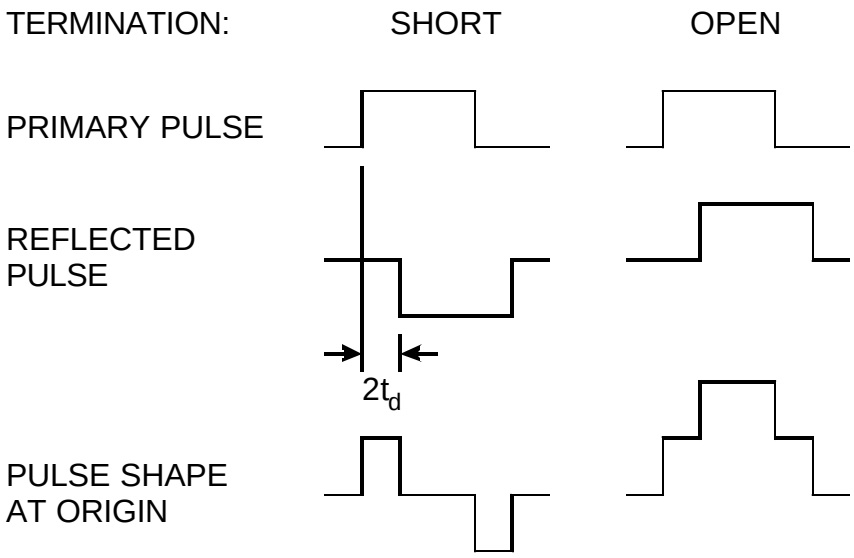
In many instances, e.g. scintillation detectors, shaping is not used to improve resolution, but to increase rate capability.

Example: delay line clipping with NaI(Tl) detector



Reminder: Reflections on Transmission Lines

Termination < Line Impedance: Reflection with opposite sign
Termination > Line Impedance: Reflection with same sign

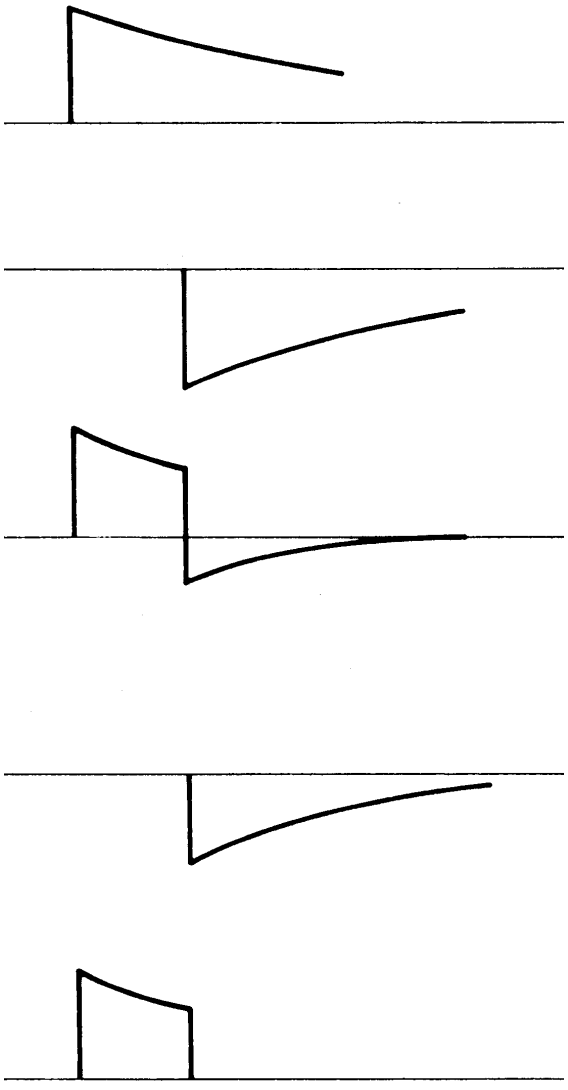


The scintillation pulse has an exponential decay.

PMT Pulse

Reflected Pulse

Sum



Eliminate undershoot by adjusting magnitude of reflected pulse

$R_T < Z_0$, but $R_T > 0$

magnitude of reflection = amplitude of detector pulse at $t = 2\ t_d$.

No undershoot at summing node ("tail compensation")

Only works perfectly for single decay time constant, but can still provide useful results when other components are much faster (or weaker).

5. Timing Measurements

Pulse height measurements discussed up to now emphasize accurate measurement of signal charge.

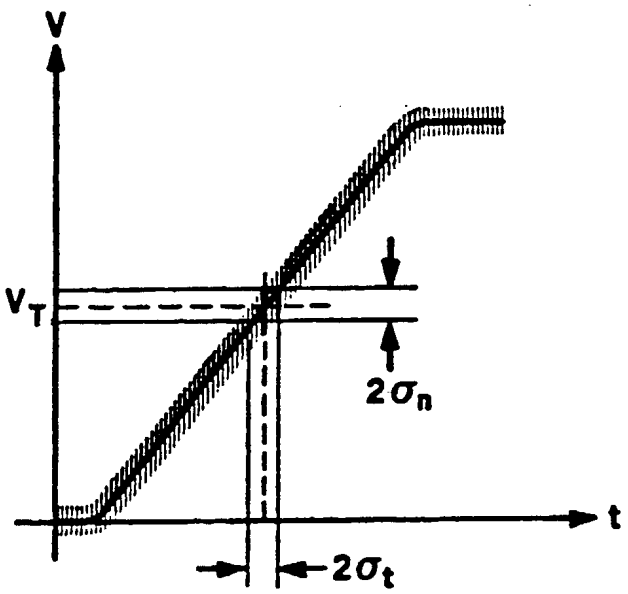
- Timing measurements optimize determination of time of occurrence.

- For timing, the figure of merit is not signal-to-noise, but slope-to-noise ratio.

Consider the leading edge of a pulse fed into a threshold discriminator (comparator).

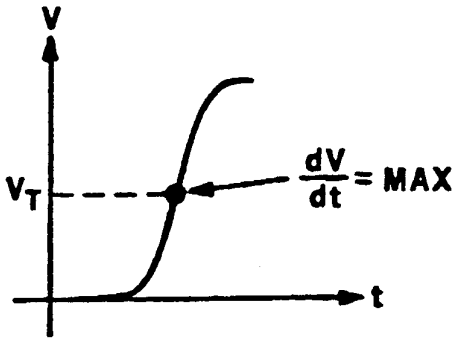
The instantaneous signal level is modulated by noise.

Þ time of threshold crossing fluctuates



$$\sigma_t = \frac{\sigma_n}{\left.\frac{dV}{dt}\right|_{V_T}}$$

Typically, the leading edge is not linear, so the optimum trigger level is the point of maximum slope.



Pulse Shaping

Consider a system whose bandwidth is determined by a single RC integrator.

The time constant of the RC low-pass filter determines the

- rise time (and hence dV/dt)
- amplifier bandwidth (and hence the noise)

Time dependence: $V_o(t) = V_0(1 - e^{-t/t})$

The rise time is commonly expressed as the interval between the points of 10% and 90% amplitude

$$t_r = 2.2\,t$$

In terms of bandwidth

$$t_r = 2.2\,t = \frac{2.2}{2p\ f_u} = \frac{0.35}{f_u}$$

Example: An oscilloscope with 100 MHz bandwidth has 3.5 ns rise time.

For a cascade of amplifiers: $t_r \approx \sqrt{t_{r1}^2 + t_{r2}^2 + \ldots + t_{rn}^2}$

Choice of Rise Time in a Timing System

Assume a detector pulse with peak amplitude V_0 and a rise time t_c passing through an amplifier chain with a rise time t_{ra}.

If the amplifier rise time is longer than the signal rise time,

$$\text{Noise} \propto \sqrt{f_u} \propto \sqrt{\frac{1}{t_{ra}}}$$

$$\frac{dV}{dt} \propto \frac{1}{t_{ra}} \propto f_u$$

increase in bandwidth ⇒ gain in dV/dt outweighs increase in noise.

In detail …

The cumulative rise time at the amplifier output (discriminator output) is

$$t_r = \sqrt{t_c^2 + t_{ra}^2}$$

The electronic noise at the amplifier output is

$$V_{no}^2 = \int e_{ni}^2 \, df = e_{ni}^2 \, \Delta f_n$$

For a single RC time constant the noise bandwidth

$$\Delta f_n = \frac{p}{2} f_u = \frac{1}{4t} = \frac{0.55}{t_{ra}}$$

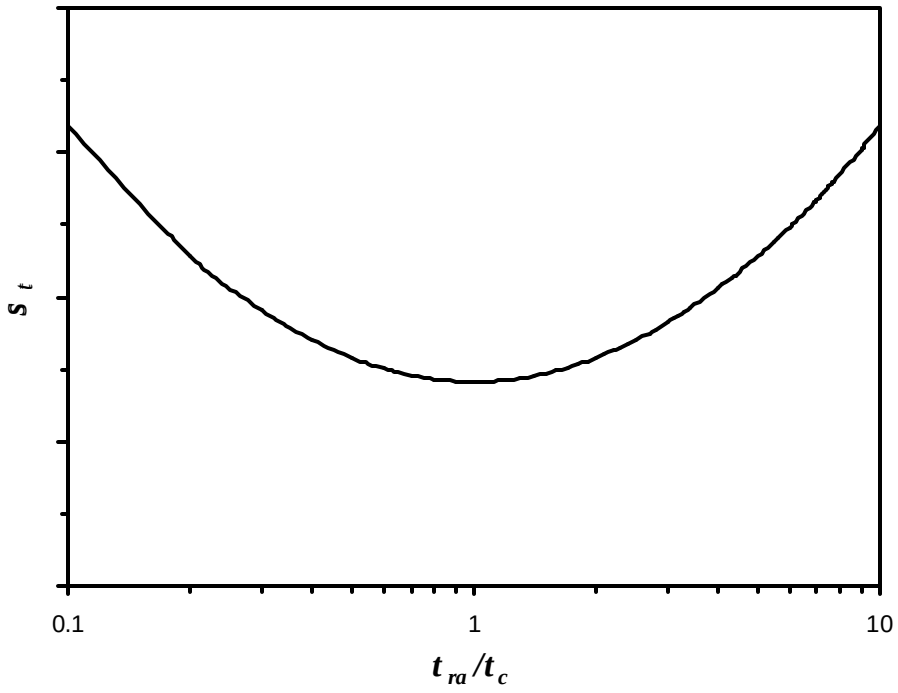
As the number of cascaded stages increases, the noise bandwidth approaches the signal bandwidth. In any case

$$\Delta f_n \propto \frac{1}{t_{ra}}$$

The timing jitter

$$s_t = \frac{V_{no}}{dV/dt} \approx \frac{V_{no}}{V_0/t_r} = \frac{1}{V_0} V_{no} t_r \propto \frac{1}{V_0} \frac{1}{\sqrt{t_{ra}}} \sqrt{t_c^2 + t_{ra}^2} = \frac{\sqrt{t_c}}{V_0} \sqrt{\frac{t_c}{t_{ra}} + \frac{t_{ra}}{t_c}}$$

The second factor assumes a minimum when the rise time of the amplifier equals the collection time of the detector $t_{ra} = t_c$.



At amplifier rise times greater than the collection time, the time resolution suffers because of rise time degradation. For smaller amplifier rise times the electronic noise dominates.

The timing resolution improves with decreasing collection time $\sqrt{t_c}$ and increasing signal amplitude V_0.

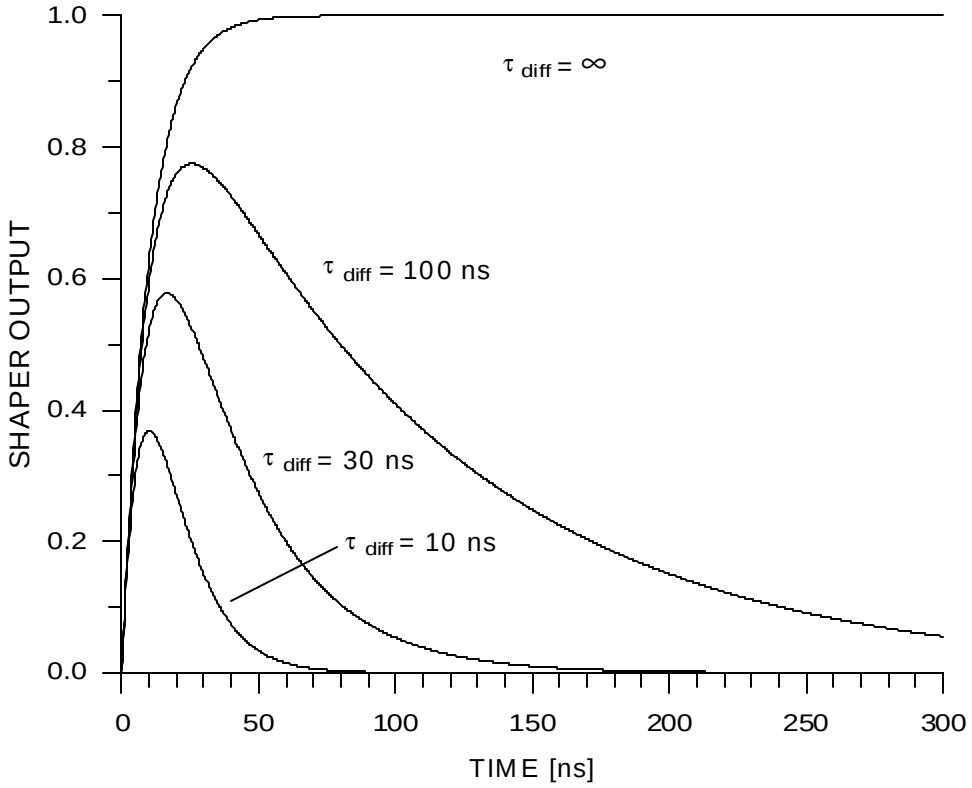
The integration time should be chosen to match the rise time.

How should the differentiation time be chosen?

As shown in the figure below, the loss in signal can be appreciable even for rather large ratios t_{diff}/t_{int}, e.g. >20% for $t_{diff}/t_{int} = 10$.

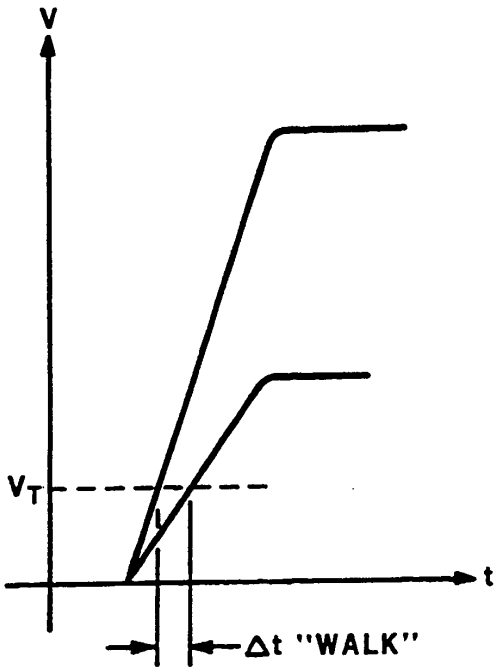
Since the time resolution improves directly with increasing peak signal amplitude, the differentiation time should be set to be as large as allowed by the required event rate.

CR-RC SHAPER
FIXED INTEGRATOR TIME CONSTANT = 10 ns
DIFFERENTIATOR TIME CONSTANT = ∞ , 100, 30 and 10 ns



Time Walk

For a fixed trigger level the time of threshold crossing depends on pulse amplitude.



- Accuracy of timing measurement limited by
  - jitter (due to noise)
  - time walk (due to amplitude variations)

If the rise time is known, “time walk” can be compensated in software event-by-event by measuring the pulse height and correcting the time measurement.

This technique fails if both amplitude and rise time vary, as is common.

In hardware, time walk can be reduced by setting the threshold to the lowest practical level, or by using amplitude compensation circuitry, e.g. constant fraction triggering.

Lowest Practical Threshold

Single RC integrator has maximum slope at $t = 0$.

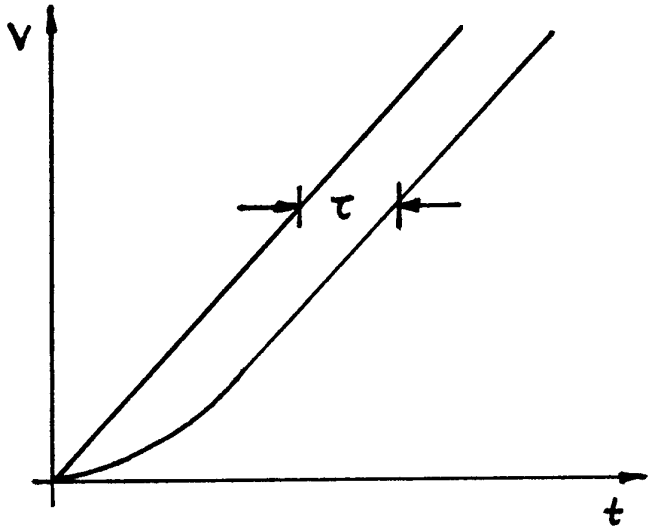
$$\frac{d}{dt}(1 - e^{-t/t}) = \frac{1}{t} e^{-t/t}$$

However, the rise time of practically all fast timing systems is determined by multiple time constants.

For small t the slope at the output of a single RC integrator is linear, so initially the pulse can be approximated by a ramp $a\ t$.

Response of the following integrator

$$V_i = a\ t \quad \rightarrow \quad V_o = a(t - t) - a\ t\ e^{-t/t}$$



$\Rightarrow$ The output is delayed by t and curvature is introduced at small t.

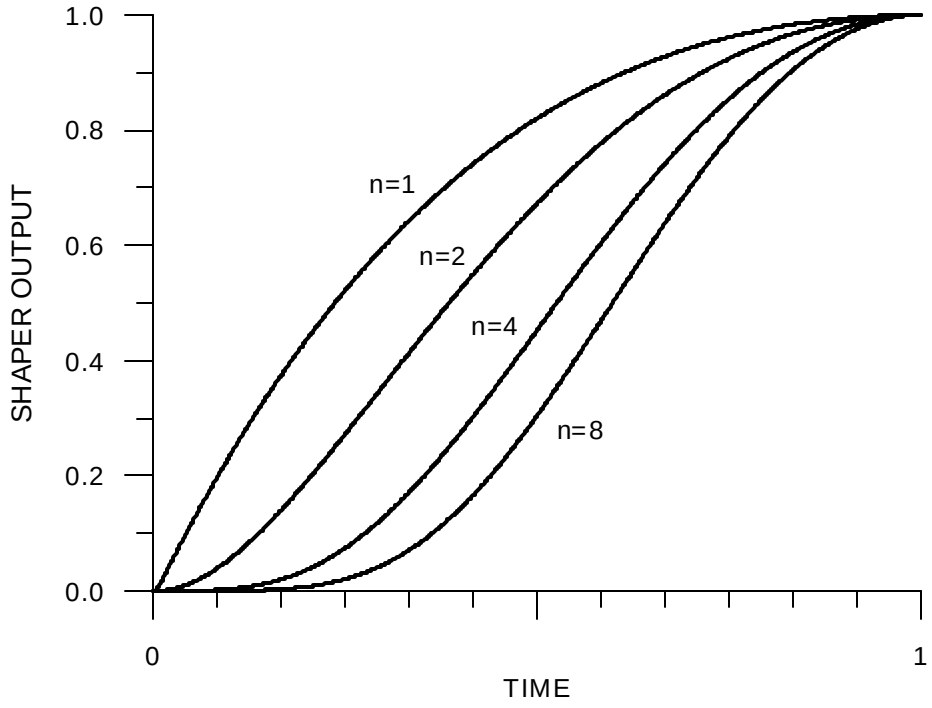
Output attains 90% of input slope after $t = 2.3\,t$.

Delay for n integrators= $n\,t$

Additional RC integrators introduce more curvature at the beginning of the pulse.

Output pulse shape for multiple RC integrators

(normalized to preserve the peaking time $t_n = t_{n=1}/n$)



Increased curvature at beginning of pulse limits the minimum threshold for good timing.

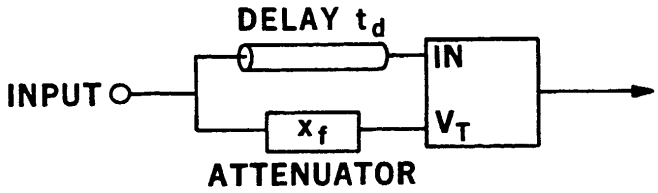
- One dominant time constant best for timing measurements

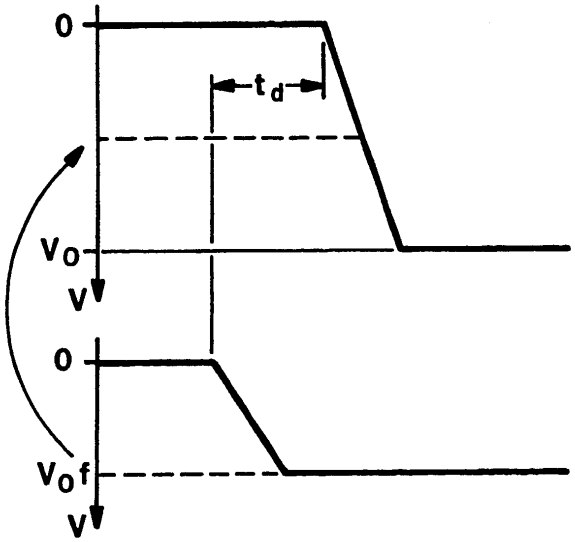
Unlike amplitude measurements, where multiple integrators are desirable to improve pulse symmetry and count rate performance.

Constant Fraction Timing

Basic Principle:

make the threshold track the signal





The threshold is derived from the signal by passing it through an attenuator $V_T = f\, V_s$.

The signal applied to the comparator input is delayed so that the transition occurs after the threshold signal has reached its maximum value $V_T = f\, V_0$.

For simplicity assume a linear leading edge

$$V(t) = \frac{t}{t_r} V_0 \quad \text{for } t \le t_r \qquad \text{and} \qquad V(t) = V_0 \quad \text{for } t > t_r$$

so the signal applied to the input is

$$V(t) = \frac{t - t_d}{t_r} V_0$$

When the input signal crosses the threshold level

$$f\, V_0 = \frac{t - t_d}{t_r} V_0$$

and the comparator fires at the time

$$t = f\, t_r + t_d \qquad (t_d > t_r)$$

at a constant fraction of the rise time independent of peak amplitude.

If the delay t_d is reduced so that the pulse transitions at the signal and threshold inputs overlap, the threshold level

$$V_T = f \frac{t}{t_r} V_0$$

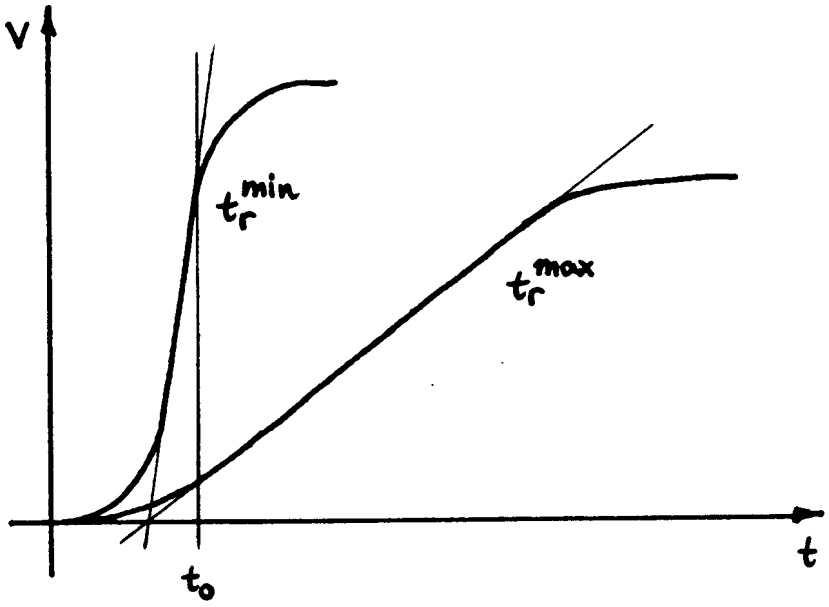
and the comparator fires at

$$f \frac{t}{t_r} V_0 = \frac{t - t_d}{t_r} V_0$$

$$t = \frac{t_d}{1 - f} \qquad (t_d < (1 - f)\, t_r)$$

independent of both amplitude and rise time (amplitude and rise-time compensation).

The circuit compensates for amplitude and rise time if pulses have a sufficiently large linear range that extrapolates to the same origin.



The condition for the delay must be met for the minimum rise time:

$$t_d \leq (1-f)\, t_{r,\min}$$

In this mode the fractional threshold V_T/V_0 varies with rise time.

For all amplitudes and rise times within the compensation range the comparator fires at the time

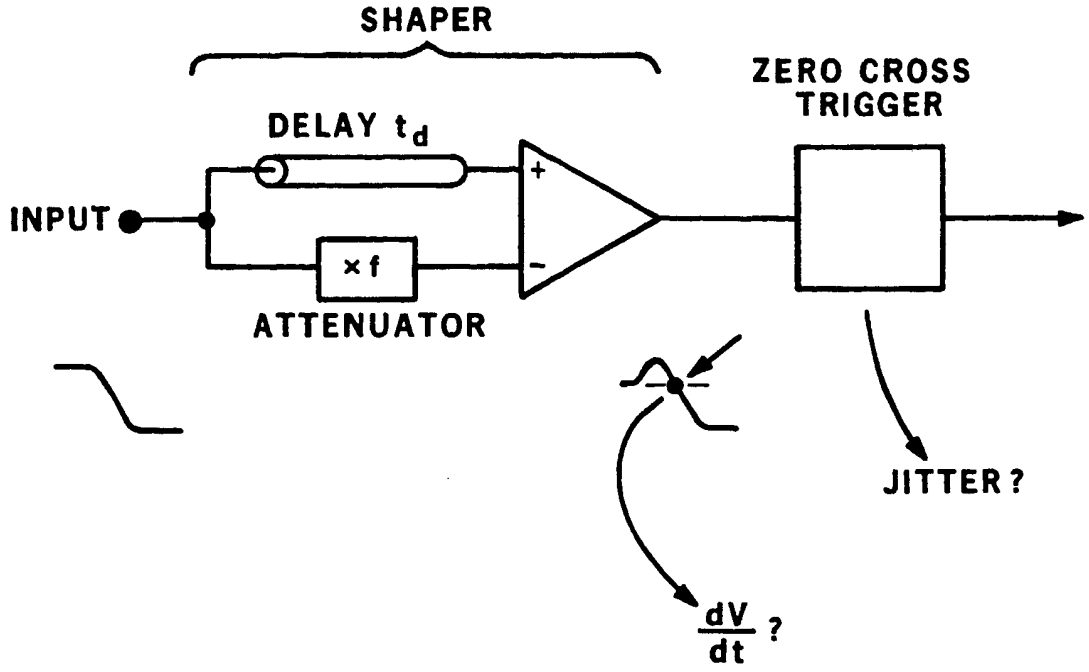
$$t_0 = \frac{t_d}{1-f}$$

Another View of Constant Fraction Discriminators

The constant fraction discriminator can be analyzed as a pulse shaper, comprising the

- delay
- attenuator
- subtraction

driving a trigger that responds to the zero crossing.

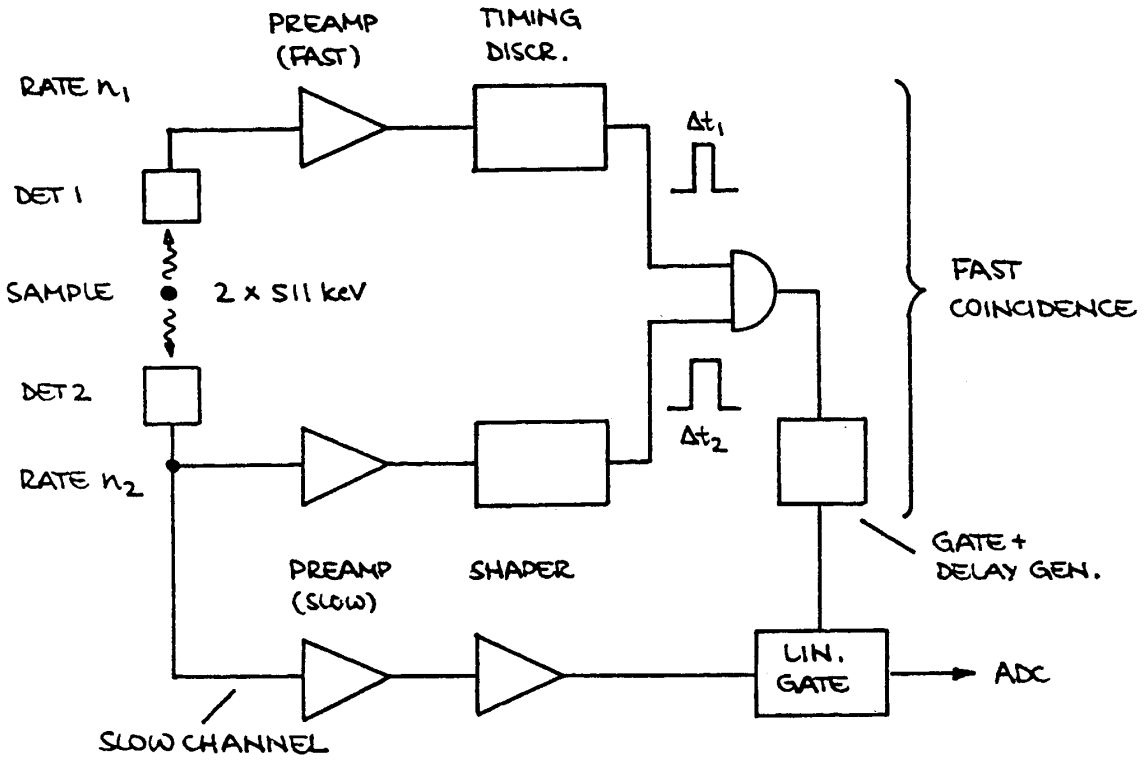


The timing jitter depends on

- the slope at the zero-crossing
  (depends on choice of f and t_d)
- the noise at the output of the shaper
  (this circuit increases the noise bandwidth)

Examples

1. *g* -*g* coincidence (as used in positron emission tomography)



Positron annihilation emits two collinear 511 keV photons.

Each detector alone will register substantial background.

Non-coincident background can be suppressed by requiring simultaneous signals from both detectors.

- Each detector feeds a fast timing channel.
- The timing pulses are combined in an AND gate (coincidence unit). The AND gate only provides an output if the two timing pulses overlap.
- The coincidence output is used to open a linear gate, that allows the energy signal to pass to the ADC.

This arrangement accommodates the contradictory requirements of timing and energy measurements. The timing channels can be fast, whereas the energy channel can use slow shaping to optimize energy resolution ("fast-slow coincidence").

Chance coincidence rate

Two random pulse sequences have some probability of coincident events.

If the event rates in the two channels are n_1 and n_2, and the timing pulse widths are Dt_1 and Dt_2, the probabality of a pulse from the first source occuring in the total coincidence window is

$$P_1 = n_1 \cdot (\Delta t_1 + \Delta t_2)$$

The coincidence is "sampled" at a rate n_2 , so the chance coincidence rate is

$$n_c = P_1 \cdot n_2$$

$$n_c = n_1 \cdot n_2 \cdot (\Delta t_1 + \Delta t_2)$$

i.e. in the arrangement shown above, the chance coincidence rate increases with the square of the source strength.

Example: $n_1 = n_2 = 10^6\ s^{-1}$

$Dt_1 = Dt_1 = 5$ ns

Þ $n_c = 10^4\ s^{-1}$

2. Nuclear Mass Spectroscopy by Time-of-Flight

Two silicon detectors

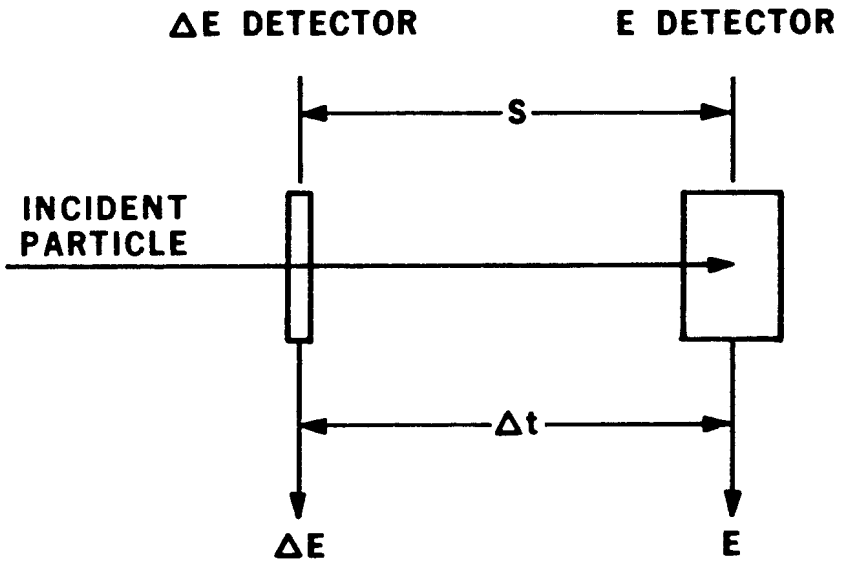
First detector thin, so that particle passes through it (transmission detector)

$\Rightarrow$ differential energy loss DE

Second detector thick enough to stop particle

$\Rightarrow$ Residual energy E

Measure time-of-flight Dt between the two detectors



$\Delta E + E \rightarrow E_{TOT}$
$\Delta E, E \rightarrow Z$
$\Delta t, E \rightarrow A$

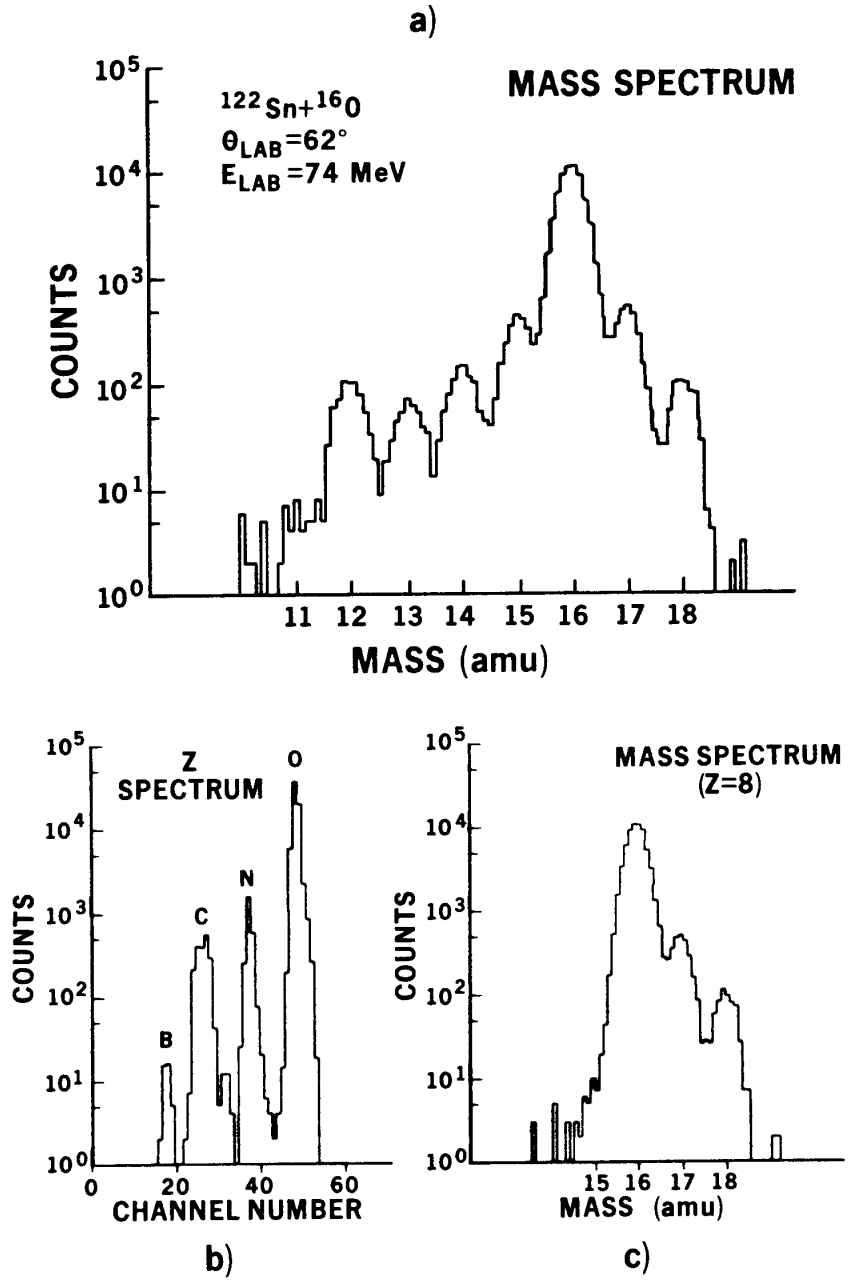
$$E_{tot} = \Delta E + E \qquad Z \propto \sqrt{\Delta E \cdot E_{tot}} \qquad A \propto E \cdot (\Delta t / s)^2$$

“Typical” Results

Example 1

Flight path 20 cm, $Dt \approx 50$ ps FWHM

$s_t = 21$ ps



(H. Spieler et al., Z. Phys. **A278** (1976) 241)

Example 2

1. DE-detector: 27 μm thick, A= 100 mm^2, <E>=1.1·10^4 V/cm

2. E-detector: 142 μm thick, A= 100 mm^2, <E>=2·10^4 V/cm

For 230 MeV ^{28}Si: DE = 50 MeV ⇒ V_s= 5.6 mV
E = 180 MeV ⇒ V_s= 106 mV

⇒ Dt = 32 ps FWHM
s_t = 14 ps

Example 3

Two transmission detectors,

each 160 μm thick, A= 320 mm^2

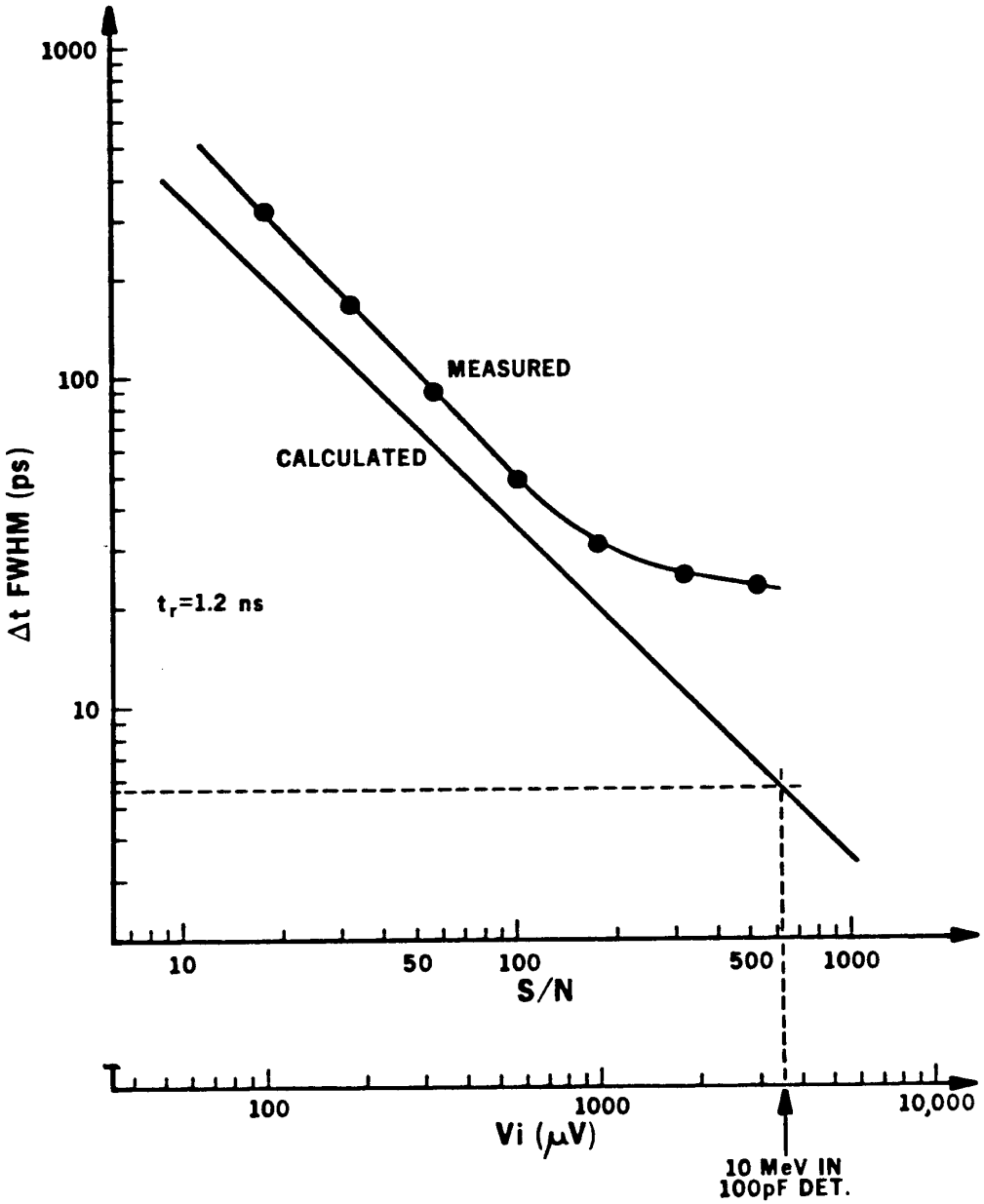
For 650 MeV/u ^{20}Ne: DE = 4.6 MeV ⇒ V_s= 800 μV

⇒ Dt = 180 ps FWHM
s_t = 77 ps

For 250 MeV/u ^{20}Ne: DE = 6.9 MeV ⇒ V_s= 1.2 mV

⇒ Dt = 120 ps FWHM
s_t = 52 ps

Fast Timing: Comparison between theory and experiment



At $S/N<100$ the measured curve lies above the calculation because the timing discriminator limited the rise time.
At high S/N the residual jitter of the time digitizer limits the resolution.

For more details on fast timing with semiconductor detectors, see H. Spieler, IEEE Trans. Nucl. Sci. **NS-29/3** (1982) 1142.

6. Digitization of Pulse Height and Time – Analog to Digital Conversion

For data storage and subsequent analysis the analog signal at the shaper output must be digitized.

Important parameters for ADCs used in detector systems:

1. Resolution
   The “granularity” of the digitized output

2. Differential Non-Linearity
   How uniform are the digitization increments?

3. Integral Non-Linearity
   Is the digital output proportional to the analog input?

4. Conversion Time
   How much time is required to convert an analog signal to a digital output?

5. Count-Rate Performance
   How quickly can a new conversion commence after completion of a prior one without introducing deleterious artifacts?

6. Stability
   Do the conversion parameters change with time?

Instrumentation ADCs used in industrial data acquisition and control systems share most of these requirements. However, detector systems place greater emphasis on differential non-linearity and count-rate performance. The latter is important, as detector signals often occur randomly, in contrast to measurement systems where signals are sampled at regular intervals.

1. Resolution

Digitization incurs approximation, as a continuous signal distribution is transformed into a discrete set of values. To reduce the additional errors (noise) introduced by digitization, the discrete digital steps must correspond to a sufficiently small analog increment.

Simplistic assumption:

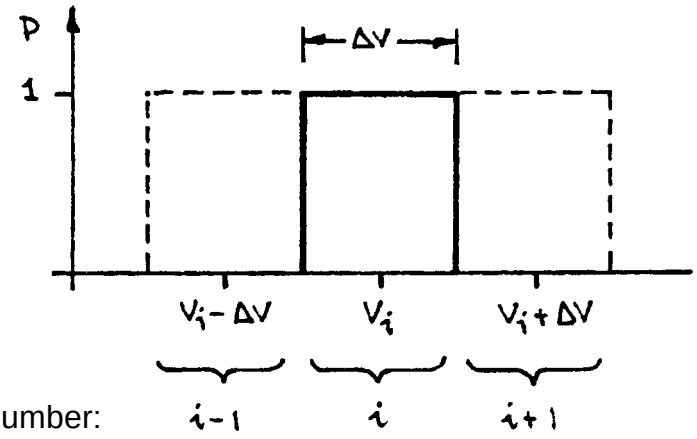
Resolution is defined by the number of output bits, e.g.

13 bits ® $\frac{\Delta V}{V} = \frac{1}{8192} = 1.2 \cdot 10^{-4}$

True Measure: Channel Profile

Plot probability vs. pulse amplitude that a pulse height corresponding to a specific output bin is actually converted to that address.

Ideal ADC:



output bin number:

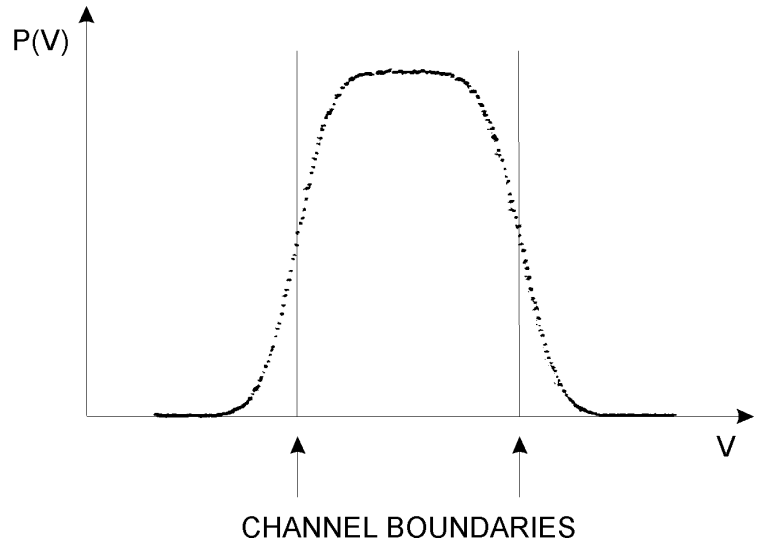
Measurement accuracy:

- If all counts of a peak fall in one bin, the resolution is ΔV.
- If the counts are distributed over several (>4 or 5) bins, peak fitting can yield a resolution of $10^{-1} – 10^{-2}\ \Delta V$, *if the distribution is known and reproducible* (not necessarily a valid assumption for an ADC).

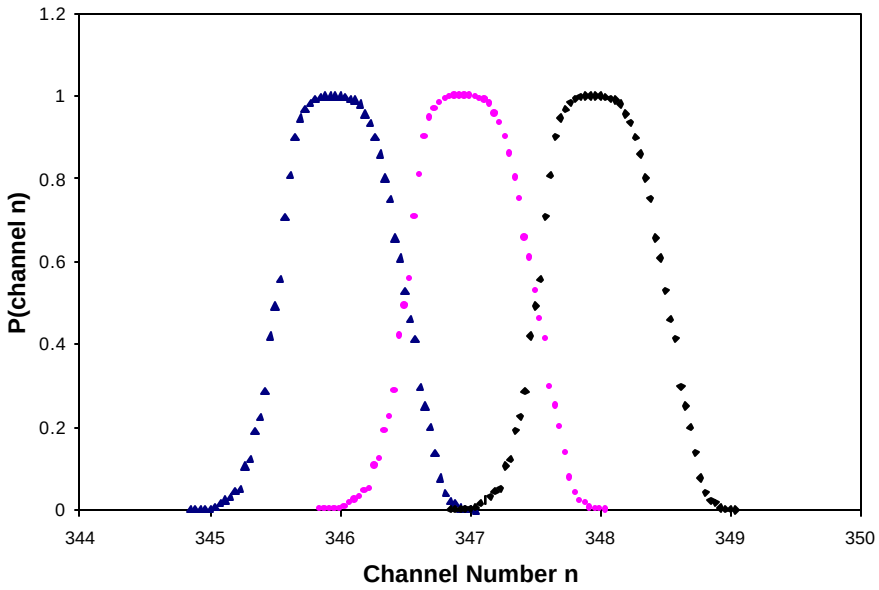
In reality, the channel profile is not rectangular as sketched above.

Electronic noise in the threshold discrimination process that determines the channel boundaries “smears” the transition from one bin to the next.

Measured channel profile (13 bit ADC)



The profiles of adjacent channels overlap

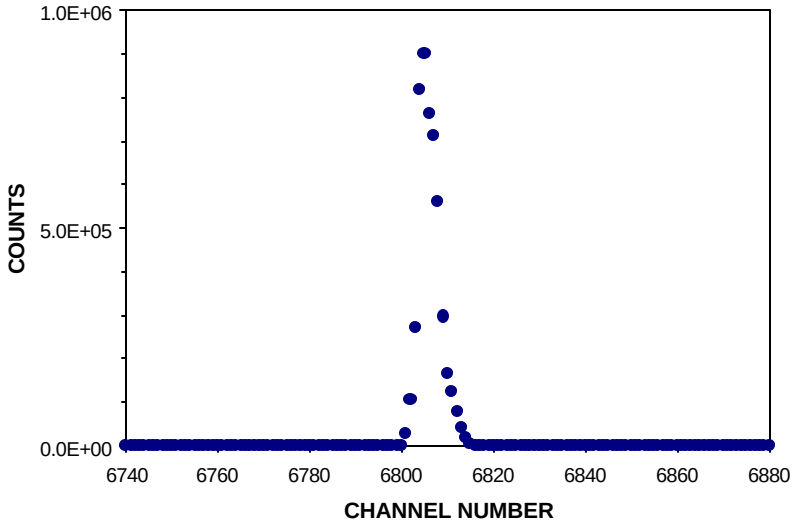


Channel profile can be checked quickly by applying the output of a precision pulser to the ADC.

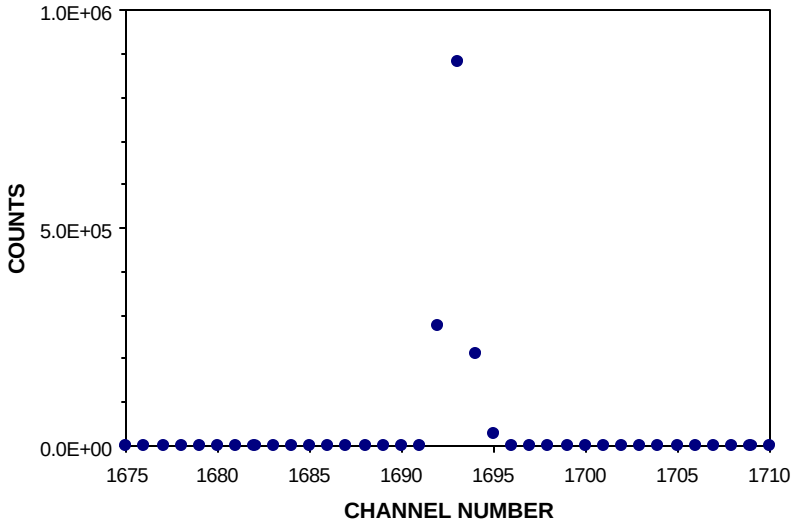
If the pulser output has very low noise, i.e. the amplitude jitter is much smaller than the voltage increment corresponding to one ADC channel or bin, all pulses will be converted to a single channel, with only a small fraction appearing in the neighbor channels.

Example of an ADC whose digital resolution is greater than its analog resolution:

8192 ch conversion range (13 bits)



2048 ch conversion range (11 bits)

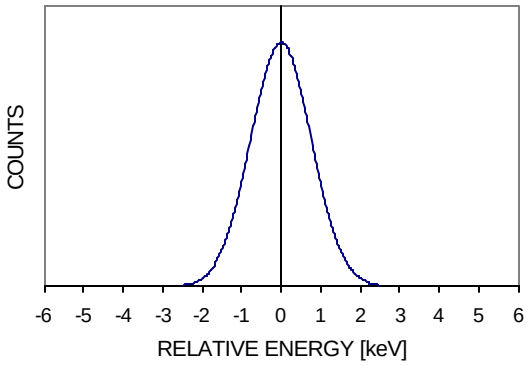


2K range provides maximum resolution – higher ranges superfluous.

How much ADC Resolution is Required?

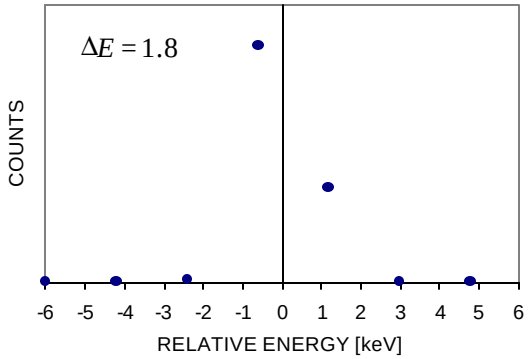
Example:

Detector resolution
1.8 keV FWHM

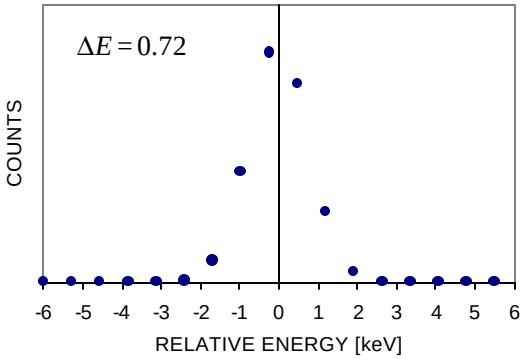


Digitized spectra for various ADC resolutions (bin widths) ΔE:

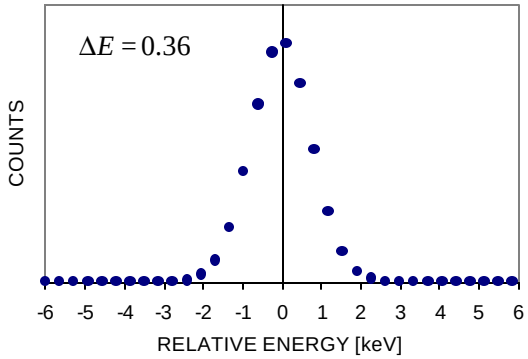
ΔE= 1.8 keV = 1 x FWHM



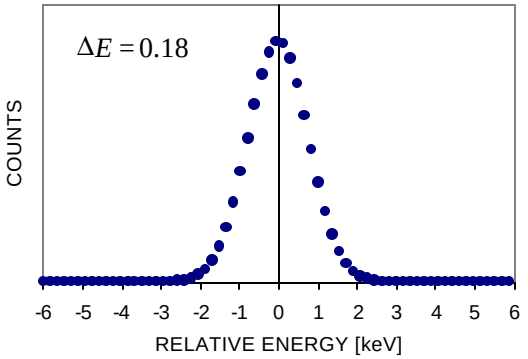
ΔE= 0.72 keV = 0.4 x FWHM



ΔE= 0.36 keV = 0.2 x FWHM



ΔE= 0.18 keV = 0.1 x FWHM



Fitting can determine centroid position to fraction of bin width even with coarse digitization, **if the line shape is known**.

Five digitizing channels within a linewidth (FWHM) allow robust peak fitting and centroid finding, even for imperfectly known line shapes and overlapping peaks.

2. Differential Non-Linearity

Differential non-linearity is a measure of the inequality of channel profiles over the range of the ADC.

Depending on the nature of the distribution, either a peak or an rms specification may be appropriate.

$$DNL = \max\left\{\frac{\Delta V(i)}{\langle \Delta V \rangle} - 1\right\}$$

or

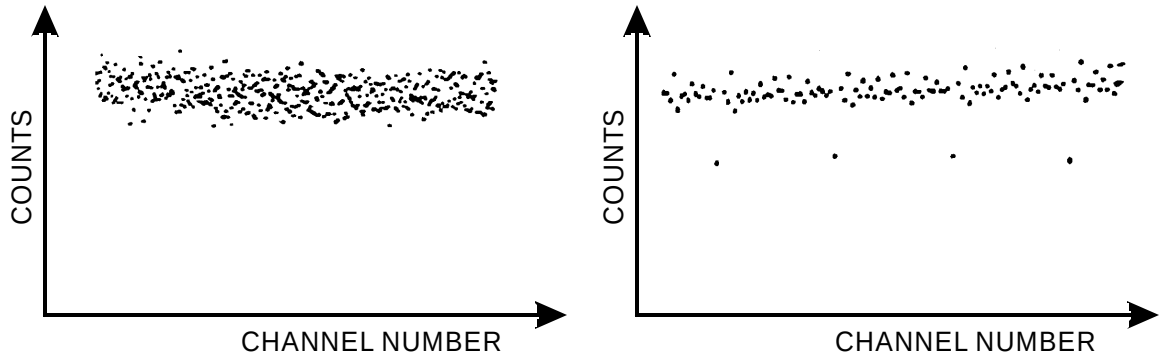
$$DNL = \text{r.m.s.}\left\{\frac{\Delta V(i)}{\langle \Delta V \rangle} - 1\right\}$$

where $\langle \Delta V \rangle$ is the average channel width and $\Delta V(i)$ is the width of an individual channel.

Differential non-linearity of < ±1% max. is typical, but state-of-the-art ADCs can achieve 10^{-3} rms, i.e. the variation is comparable to the statistical fluctuation for 10^6 random counts.

Note: Instrumentation ADCs are often specified with an accuracy of ±0.5 LSB (least significant bit), so the differential non-linearity may be 50% or more.

Typical differential non-linearity patterns (“white” input spectrum).

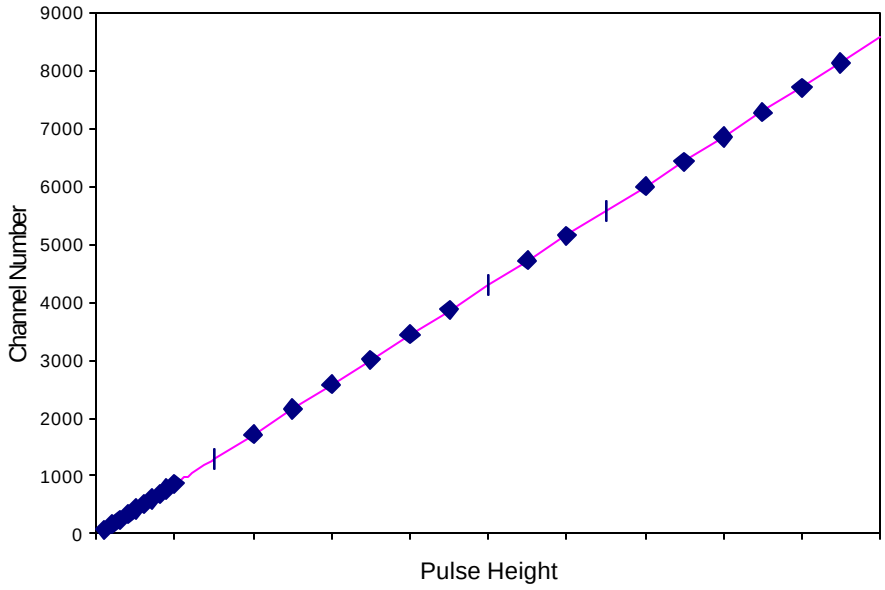


An ideal ADC would show an equal number of counts in each bin.

The spectrum to the left shows a random pattern, but note the multiple periodicities visible in the right hand spectrum.

3. Integral Non-Linearity

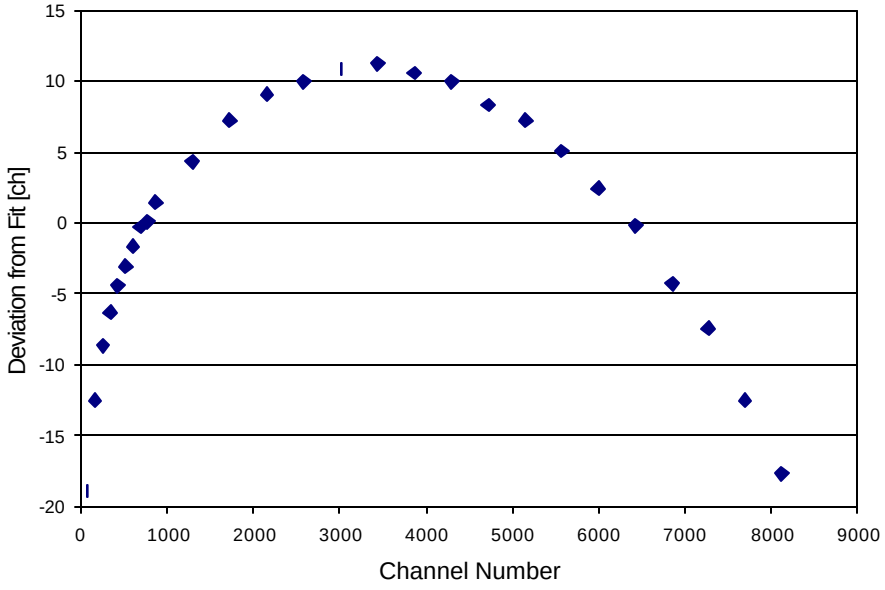
Integral non-linearity measures the deviation from proportionality of the measured amplitude to the input signal level.



The dots are measured values and the line is a fit to the data.

This plot is not very useful if the deviations from linearity are small.

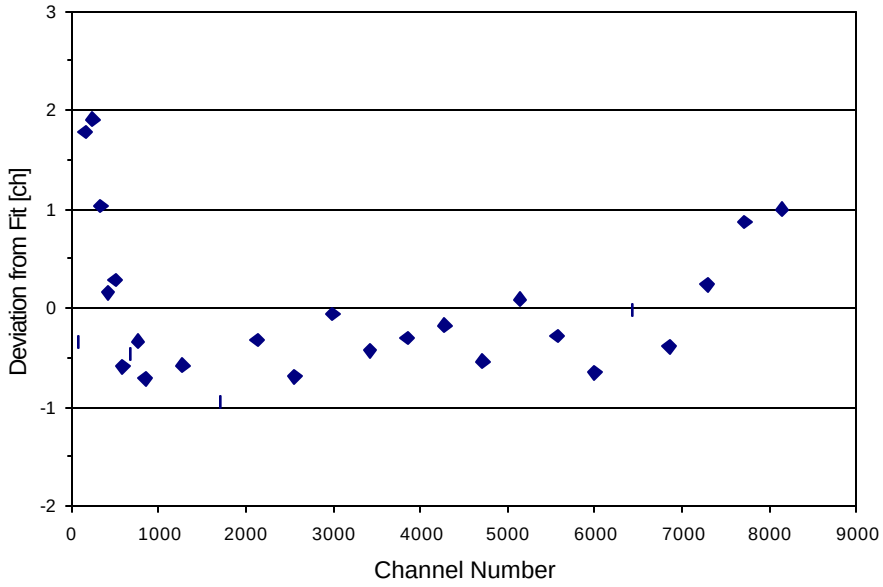
Plotting the deviations of the measured points from the fit yields:



The linearity of an ADC can depend on the input pulse shape and duration, due to bandwidth limitations in the circuitry.

The differential non-linearity shown above was measured with a 400 ns wide input pulse.

Increasing the pulse width to 3 µs improved the result significantly:



4. Conversion Time

During the acquisition of a signal the system cannot accept a subsequent signal (“dead time”)

Dead Time =

signal acquisition time	→	time-to-peak + const.
+ conversion time	→	can depend on pulse height
+ readout time to memory	→	depends on speed of data transmission and buffer memory access - can be large in computer-based systems

Dead time affects measurements of yields or reaction cross-sections. Unless the event rate << 1/(dead time), it is necessary to measure the dead time, e.g. with a reference pulser fed simultaneously into the spectrum.

The total number of reference pulses issued during the measurement is determiend by a scaler and compared with the number of pulses recorded in the spectrum.

Does a pulse-height dependent dead time mean that the correction is a function of pulse height?

Usually not. If events in different part of the spectrum are not correlated in time, i.e. random, they are all subject to the same average dead time (although this average will depend on the spectral distribution).

- Caution with correlated events!
  Example: Decay chains, where lifetime is < dead time.
  The daughter decay will be lost systematically.

5. Count Rate Effects

Problems are usually due to internal baseline shifts with event rate or undershoots following a pulse.

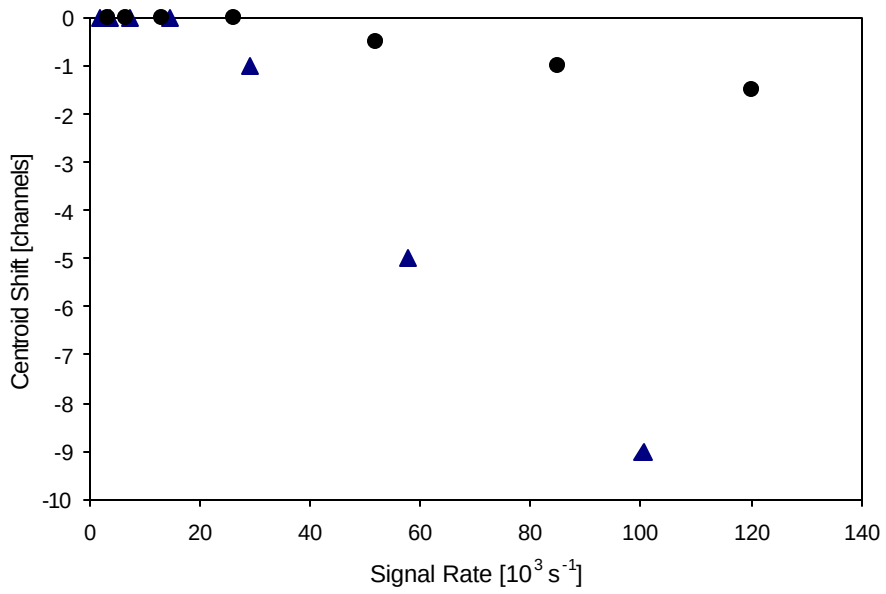
If signals occur at constant intervals, the effect of an undershoot will always be the same.

However, in a random sequence of pulses, the effect will vary from pulse to pulse.

- spectral broadening

Baseline shifts tend to manifest themselves as a systematic shift in centroid position with event rate.

Centroid shifts for two 13 bit ADCs vs. random rate:



6. Stability

   Stability vs. temperature is usually adequate with modern electronics in a laboratory environment.

- Note that temperature changes within a module are typically much smaller than ambient.

  However: Highly precise or long-term measurements require spectrum stabilization to compensate for changes in gain and baseline of the overall system.

  Technique: Using precision pulsers place a reference peak at both the low and high end of the spectrum.

  (Pk. Pos. 2) – (Pk. Pos. 1) $\rightarrow$ Gain, … then

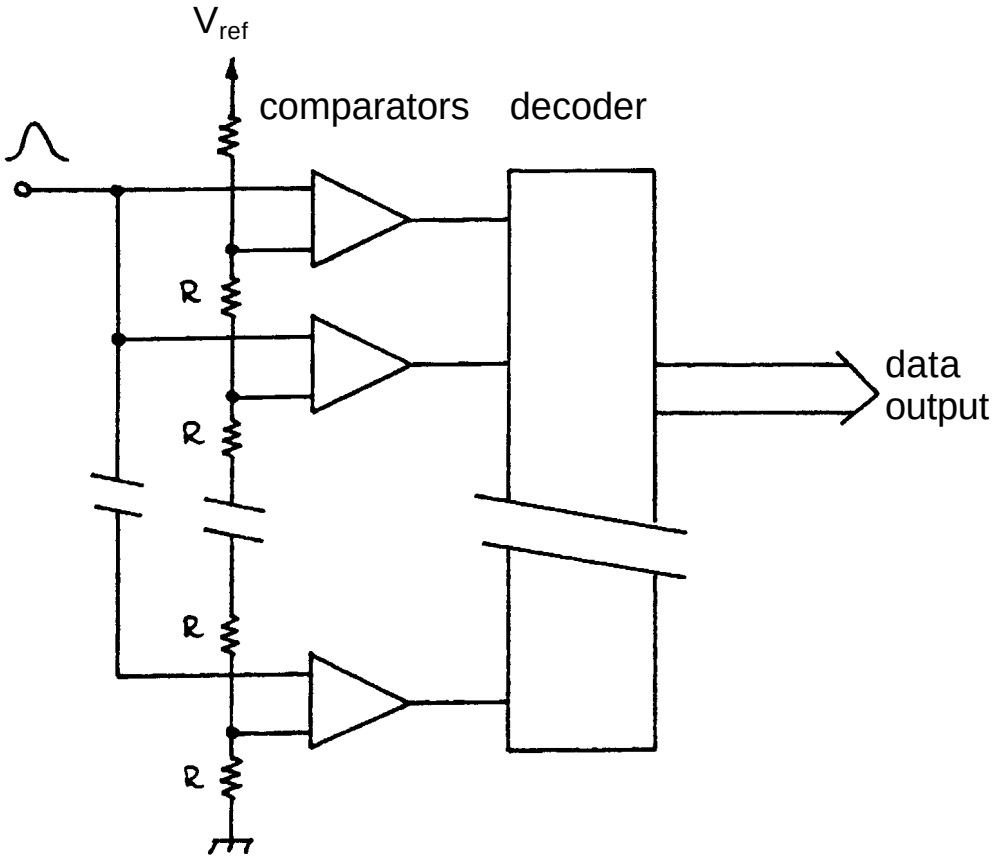
  (Pk. Pos. 1) or (Pk. Pos. 2) $\rightarrow$ Offset

  Traditional Implementation: Hardware,
  spectrum stabilizer module

  Today, it is more convenient to determine the corrections in software. These can be applied to calibration corrections or used to derive an electrical signal that is applied to the hardware (simplest and best in the ADC).

Analog to Digital Conversion Techniques

1. Flash ADC



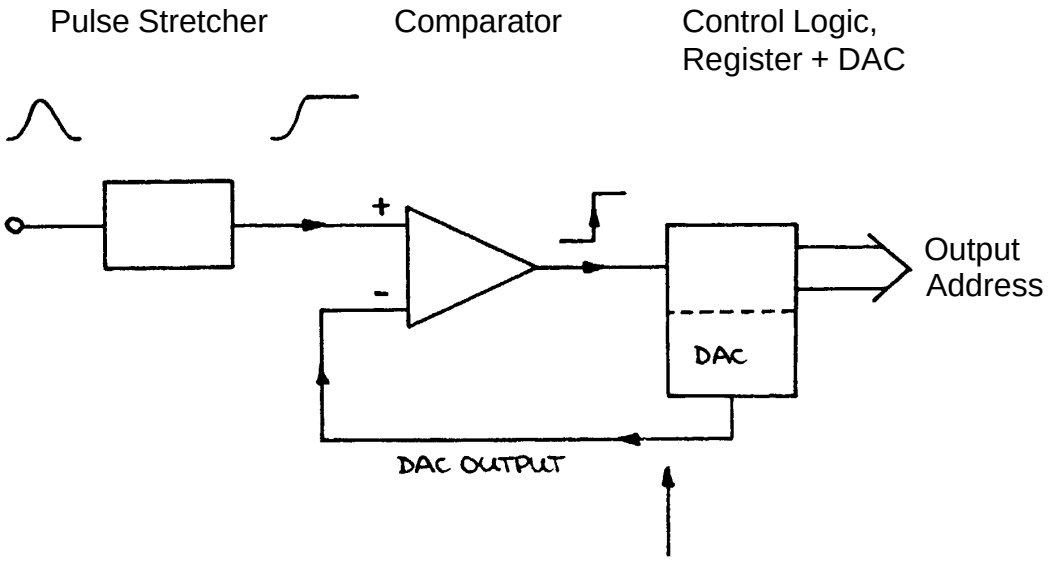
The input signal is applied to *n* comparators in parallel. The switching thresholds are set by a resistor chain, such that the voltage difference between individual taps is equal to the desired measurement resolution.

2^n comparators for *n* bits (8 bit resolution requires 256 comparators)

Feasible in monolithic ICs since the absolute value of the resistors in the reference divider chain is not critical, only the relative matching.

Advantage: short conversion time (<10 ns available)
Drawbacks: limited accuracy (many comparators)
power consumption
Differential non-linearity ~ 1%
High input capacitance (speed is often limited by the analog driver feeding the input)

2. Successive Approximation ADC



Sequentially add levels proportional to 2^n, 2^{n-1}, … 2^0 and set corresponding bit if the comparator output is high (DAC output < pulse height)

n conversion steps yield 2^n channels,
i.e. 8K channels require 13 steps

Advantages: speed (~ μs)
high resolution
ICs (monolithic + hybrid) available

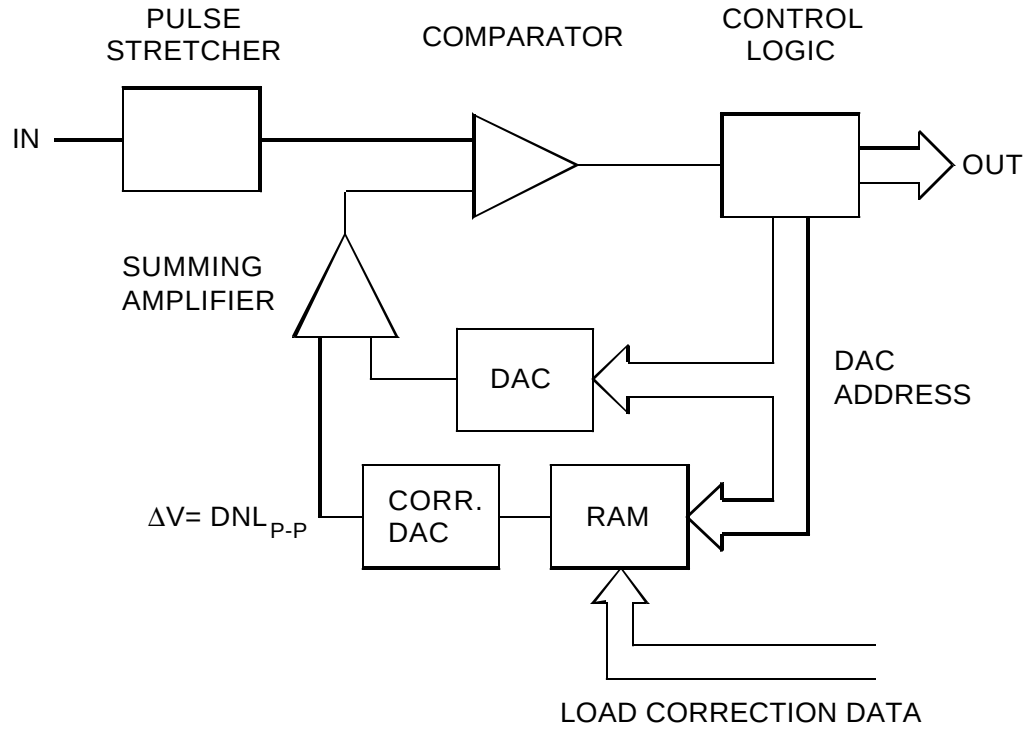
Drawback: Differential non-linearity (typ. 10 – 20%)

Reason: Resistors that set DAC output must be extremely accurate.

For DNL < 1% the resistor determining the 2^{12} level in an 8K ADC must be accurate to < $2.4 \cdot 10^{-6}$.

DNL can be corrected by various techniques:

- averaging over many channel profiles for a given pulse amplitude (“sliding scale” or “Gatti principle”)
- correction DAC (“brute force” application of IC technology)



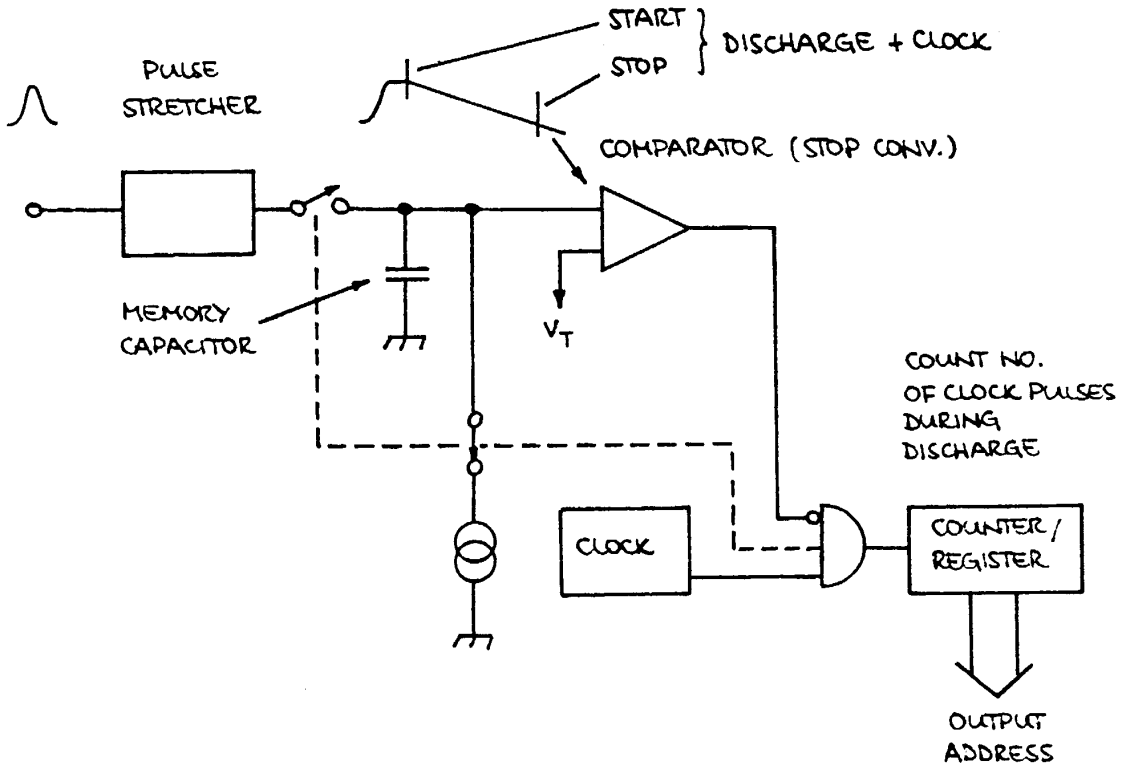
The primary DAC output is adjusted by the output of a correction DAC to reduce differential non-linearity.

Correction data are derived from a measurement of DNL. Corrections for each bit are loaded into the RAM, which acts as a look-up table to provide the appropriate value to the correction DAC for each bit of the main DAC.

The range of the correction DAC must exceed the peak-to-peak differential non-linearity.

If the correction DAC has N bits, the maximum DNL is reduced by $1/2^{(N-1)}$ (if deviations are symmetrical).

3. Wilkinson ADC



The peak signal amplitude is acquired by a pulse stretcher and transferred to a memory capacitor. Then, simultaneously,

1. the capacitor is disconnected from the stretcher,
2. a current source is switched to linearly discharge the capacitor,
3. a counter is enabled to determine the number of clock pulses until the voltage on the capacitor reaches the baseline.

Advantage: excellent differential linearity
(continuous conversion process)

Drawbacks: slow – conversion time = $n \cdot T_{clock}$
(n = channel number $\propto$ pulse height)
T_{clock} = 10 ns $\rightarrow T_{conv}$ = 82 µs for 13 bits
Clock frequencies of 100 MHz typical, but >400 MHz possible with excellent performance

"Standard" technique for high-resolution spectroscopy.

Hybrid Analog-to-Digital Converters

Conversion techniques can be combined to obtain high resolution and short conversion time.

1. Flash + Successive Approximation or
   Flash + Wilkinson (Ramp Run-Down)

   Utilize fast flash ADC for coarse conversion
   (*e.g.* 8 out of 13 bits)

   Successive approximation or Wilkinson converter to provide fine resolution. Limited range, so short conversion time:

   256 ch with 100 MHz clock $\Rightarrow$ 2.6 μs

   Results: 13 bit conversion in < 4 μs
   with excellent integral and differential linearity

2. Flash ADCs with Sub-Ranging

   Not all applications require constant absolute resolution over the full range. Sometimes only *relative* resolution must be maintained, especially in systems with a very large dynamic range.

   Precision binary divider at input to determine coarse range + fast flash ADC for fine digitization.

   Example: Fast digitizer that fits in phototube base.
   Designed at FNAL.

   17 to 18 bit dynamic range
   Digital floating point output
   (4 bit exponent, 8+1 bit mantissa)
   16 ns conversion time

Time Digitizers

1. Counter

Simplest arrangement.

Count clock pulses between start and stop.

Limitation: Speed of counter

Current technology limits speed of counter system to about 1 GHz

➧ Δt = 1 ns

Multi-hit capability

2. Analog Ramp

Commonly used in high-resolution digitizers (Δt = 10 ps)

Principle: charge capacitor through switchable current source

Start pulse: turn on current source

Stop pulse: turn off current source

➧ Voltage on storage capacitor

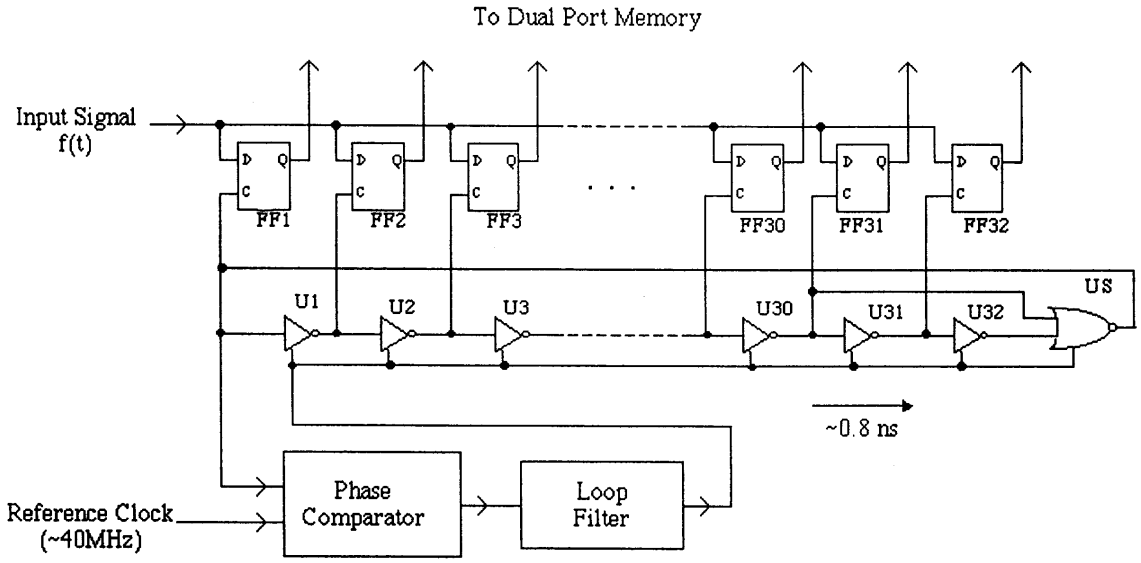
Use Wilkinson ADC with smaller discharge current to digitize voltage.

Drawbacks: No multi-hit capability
Deadtime

3. Digitizers with Clock Interpolation

Most experiments in HEP require multi-hit capability, no deadtime

Commonly used technique for time digitization (Y. Arai, KEK)

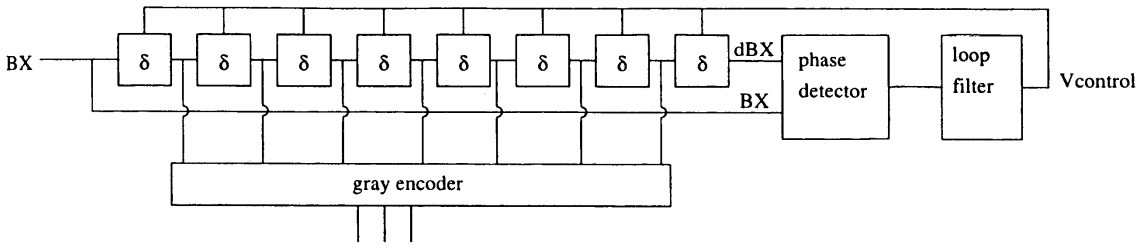


- Digitiging an external signal at a period of internal gate delay of ~0.8ns.
- High Stability with a Phase Locked Loop.
- Long time range (>3 us) & No deadtime by a Dual Port Memory.
- High precision, High Density & Low Cost LSI.

Patent Pending
- S63-067314 (JP)
- H3-133169 (JP)
- H6-69507 (JP)
- 95300652.5 (EU)

Clock period interpolated by inverter delays (U1, U2, …).
Delay can be fine-tuned by adjusting operating point of inverters.

Delays stabilized by delay-locked loop

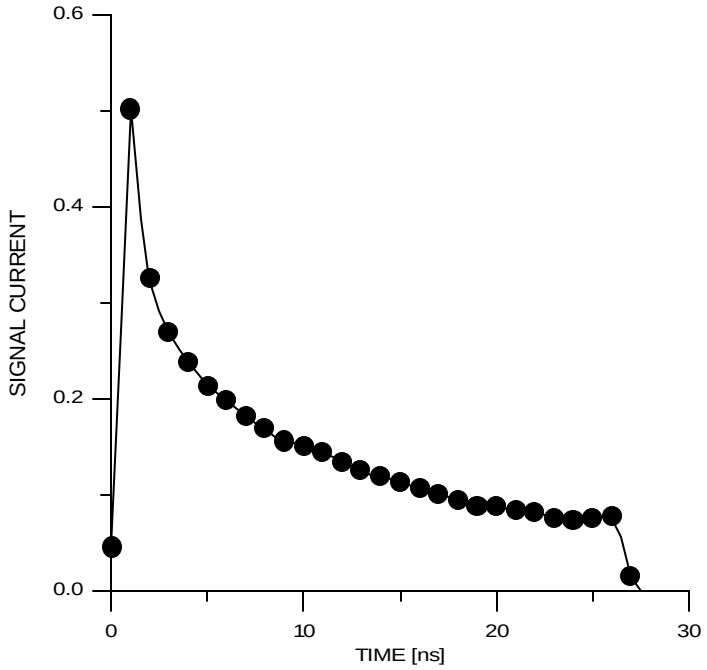


Devices with 250 ps resolution fabricated and tested.

see Y. Arai et al., IEEE Trans. Nucl. Sci. **NS-45/3** (1998) 735-739
and references therein.

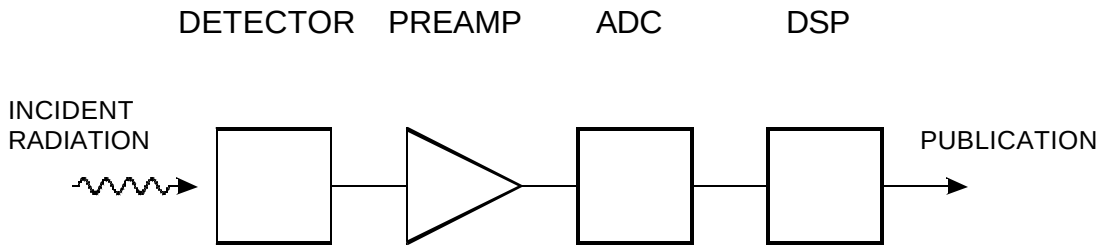
6. Digital Signal Processing

Sample detector signal with fast digitizer to reconstruct pulse:



Then use digital signal processor to perform mathematical operations for desired pulse shaping.

Block Diagram

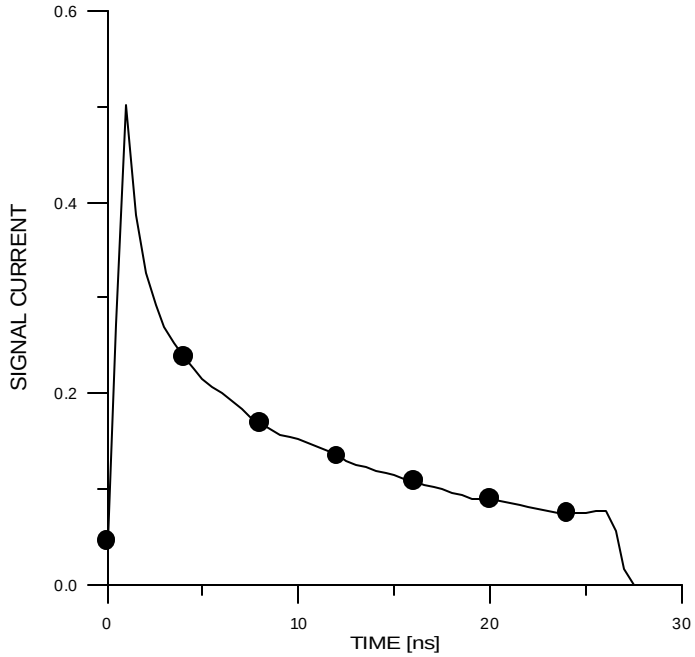


DSP allows great flexibility in implementing filtering functions

However: increased circuit complexity
increased demands on ADC,
compared to traditional shaping.

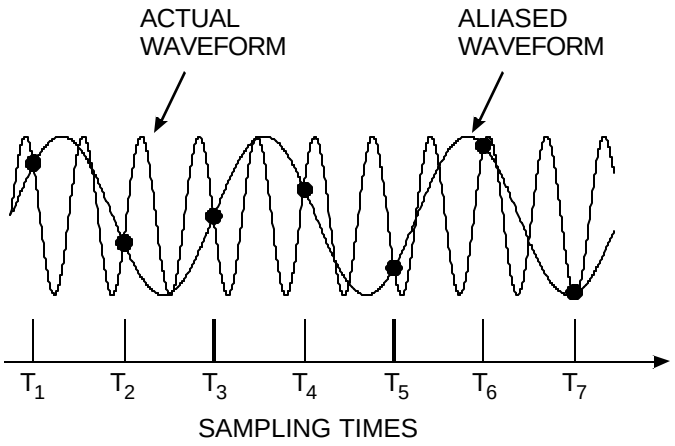
Important to choose sample interval sufficiently small to capture pulse structure.

Sampling interval of 4 ns misses initial peak.



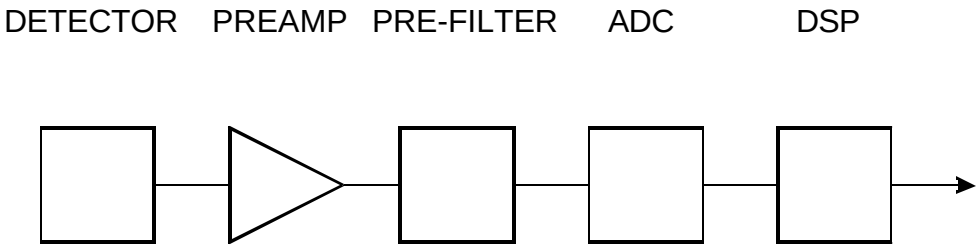
ADC must be capable of digitizing at more than twice the rate of the highest frequency component in the signal (Nyquist criterion).

With too low a sampling rate high frequency components will be "aliased" to lower frequencies:



Applies to any form of sampling (time waveform, image, ...)

Þ Fast ADC required + Pre-Filter to limit signal bandwidth



- Dynamic range requirements for ADC may be more severe than in analog filtered system (depending on pulse shape and pre-filter).

- Digitization introduces additional noise ("quantization noise")

  If one bit corresponds to an amplitude interval D, the quantization noise

$$s_v^2 = \int_{-\Delta/2}^{\Delta/2} \frac{v^2}{\Delta} dv = \frac{\Delta^2}{12}$$

  (differential non-linearity introduces quasi-random noise)

- Electronics preceding ADC and front-end of ADC must exhibit same precision as analog system, i.e.

  baseline and other pulse-to-pulse amplitude fluctuations less than order Q_n/10, i.e. typically 10^{-4} in high-resolution systems.

  For 10 V FS at the ADC input in a high-resolution gamma-ray detector system, this corresponds to < 1 mV.

Þ ADC must provide high performance at short conversion times

Today this is technically feasible for some applications, e.g. detectors with moderate to long collection times (γ and x-ray detectors).

Systems commercially available.

Benefits of digital signal processing:

- Flexibility in implementing filter functions
- Filters possible that are impractical in hardware
- Simple to change filter parameters
- Tail cancellation and pile-up rejection easily incorporated
- Adaptive filtering can be used to compensate for pulse shape variations.

Where is digital signal processing appropriate?

- Systems highly optimized for
  - Resolution
  - High counting rates
- Variable detector pulse shapes

Where is analog signal processing best (most efficient)?

- Fast shaping
- Systems not sensitive to pulse shape (fixed shaper constants)
- High density systems that require
  - small circuit area
  - low power

Both types of systems require careful analog design.

Progress in fast ADCs (precision, reduced power) will expand range of DSP applications

7. Why Things Don't Work

or

Why *S/N* Theory Often Seems to be Irrelevant

Throughout the previous lectures it was assumed that the only sources of noise were

- random
- known
- in the detector, preamplifier, or associated components

In practice, the detector system will pick up spurious signals that are

- not random,
- but not correlated with the signal,

so with reference to the signal they are quasi-random.

- Baseline fluctuations superimposed on the desired signal
- Increased detection threshold, Degradation of resolution

Important to distinguish between

- pickup of spurious signals, either from local or remote sources (clocks, digital circuitry, readout lines),

  and

- self-oscillation
  (circuit provides feedback path that causes sustained oscillation due to a portion of the output reaching the input)

Useful references:

H.W. Ott, *Noise Reduction Techniques in Electronic Systems*
Wiley, 1976, ISBN 0-471-65726-3, TK7867.5.087
H. Johnson and M. Graham, *High-Speed Digital Design*
Prentice Hall PTR, 1993, ISBN 0-13-395724-1, TK7868.D5J635

External Pickup is often the cause, but many problems are due to poor work practices or inappropriate equipment

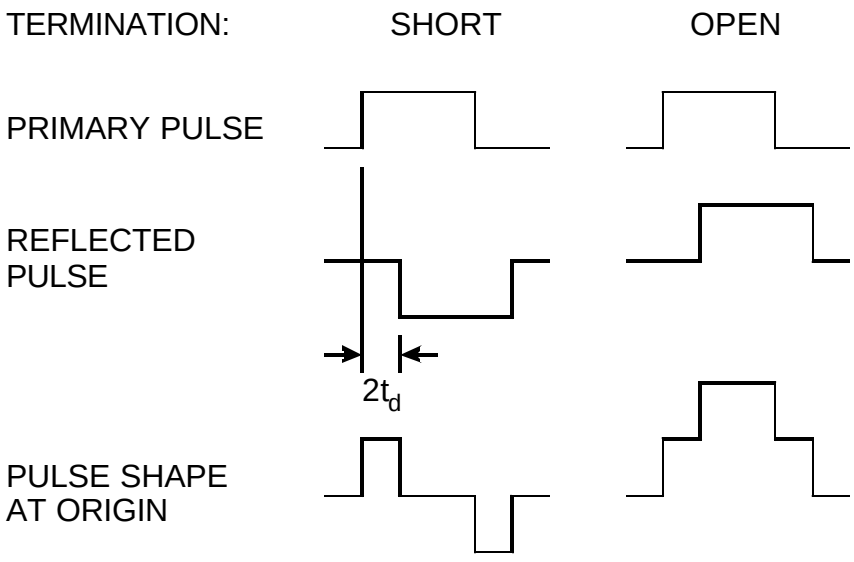
1. Termination of Cables

Signals are transmitted from one unit to another through transmission lines, often coaxial cables or ribbon cables.

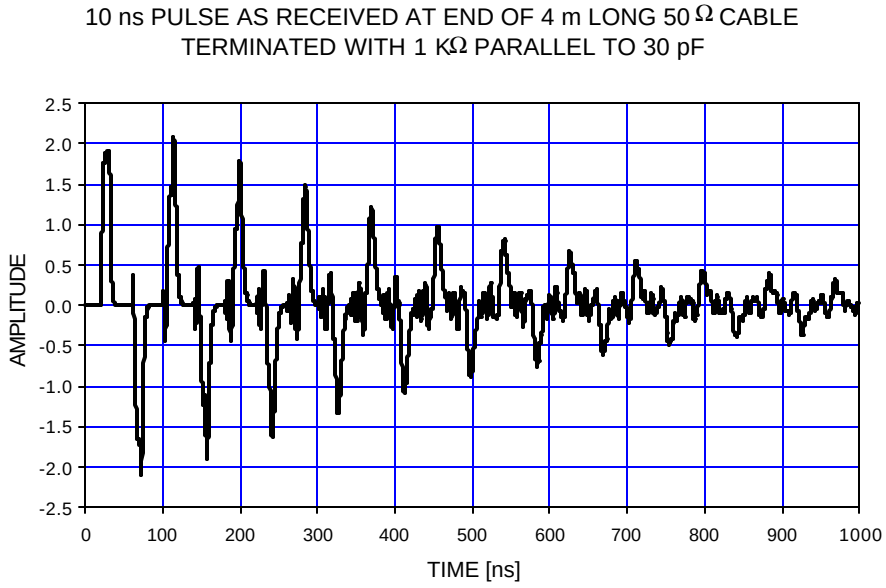
When transmission lines are not terminated with their characteristic impedance, the signals are reflected.

Reflections on Transmission Lines

Termination < Line Impedance: Reflection with opposite sign
Termination > Line Impedance: Reflection with same sign



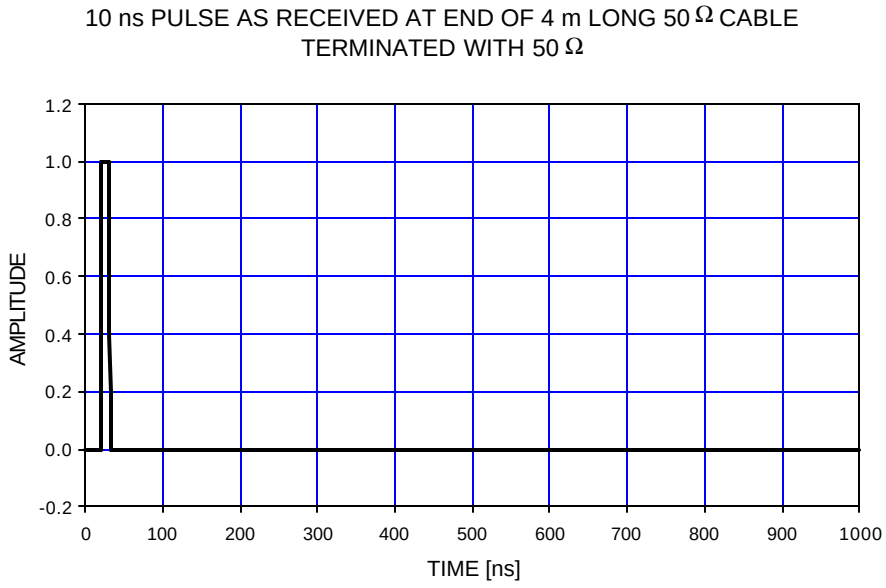
Example:



1 K in parallel with 30 pF is a typical input impedance for a shaping amplifier.

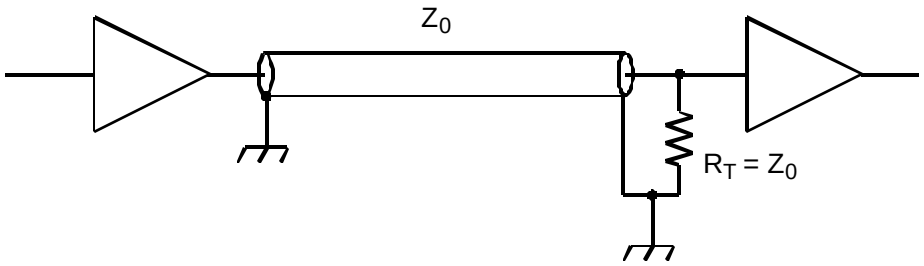
If feeding a counter, each signal pulse will be registered multiple times, depending on the threshold setting.

When the cable is properly terminated, the reflections disappear.

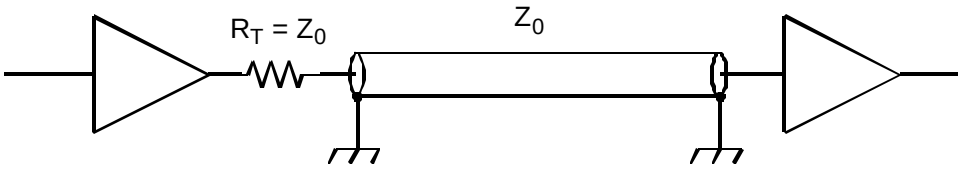


Two methods of terminating cables:

1. Parallel termination at receiving end



2. Series termination at sending end



In this configuration the original pulse is reflected at the receiving end, but the reflection is absorbed at the sending end, so it doesn't reappear at the receiver.

The series resistor feeding the transmission line forms a voltage divider that attenuates the pulse amplitude by a factor of 2.

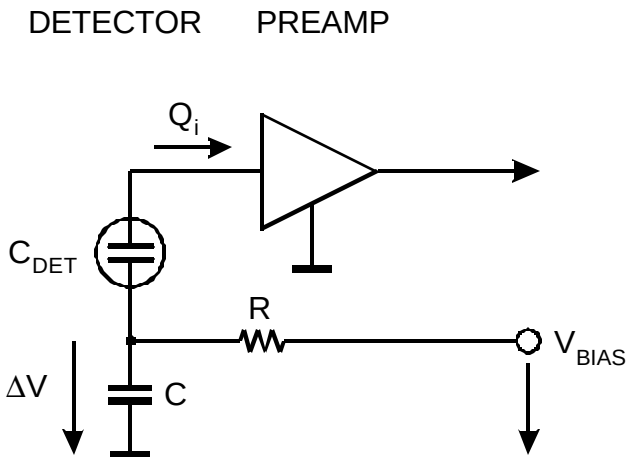
However, since the high impedance at the end of the cable causes a reflected pulse with the same amplitude, the pulse amplitude is doubled, so it is the same as the original.

Terminations are never perfect, especially at high frequencies (>10 ... 100 MHz), so in critical applications one can use both series and parallel termination. However, this does incur a 50% reduction in pulse amplitude.

In the >µs regime, amplifier inputs are usually high impedance, whereas timing amplifiers tend to be internally terminated (check!).

Noisy Detector Bias Supplies

The detector is the most sensitive node in the system.



Any disturbance ΔV on the detector bias line will induce charge in the input circuit.

$$\Delta Q = C_{\mathrm{det}} \Delta V$$

ΔV= 100 μV and 10 pF detector capacitance yield
ΔQ = 1 fC – about 6000 electrons or 20 keV (Si).

Especially when the detector bias is low (<100V), it is tempting to use a general laboratory power supply.

Frequently, power supplies are very noisy – especially old units.
The RC circuits in the bias line provide some filtering, but usually not enough for a typical power supply

Beware of switching power supplies. Well-designed switching regulators can be very clean, but most switchers are very noisy.

Spikes on the output can be quite large, but short, so that the rms noise specification may appear adequate.

Use very low noise power supplies.

Common Types of Interference

1. Light Pick-Up

Critical systems:

- Photomultiplier tubes
- Semiconductor detectors
  (all semiconductor detectors are photodiodes)

Sources

- Room lighting (Light Leaks)
- Vacuum gauges

Interference is correlated with the power line frequency
(60 Hz here, 50 Hz in Europe, Japan)

Pickup from incandescent lamps has twice the line frequency
(light intensity $\propto$ voltage squared)

Diagnostics:

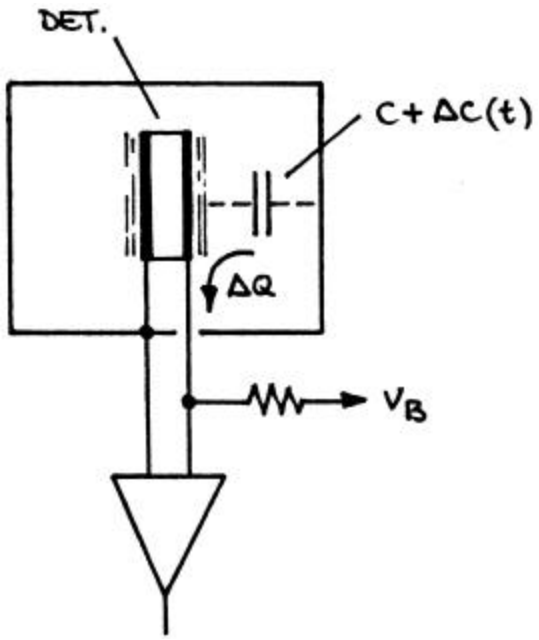
a) inspect signal output with oscilloscope set to trigger mode “line”. Look for stationary structure on baseline

Analog oscilloscope better than digital.

b) switch off light

c) cover system with black cloth (preferably felt, or very densely woven – check if you can see through it)

2. Microphonics



If the electrode at potential V_B vibrates with respect to the enclosure, the stray capacitance C is modulated by $\Delta C(t)$, inducing a charge

$$\Delta Q(t) = V_B \Delta C(t)$$

in the detector signal circuit.

Typically, vibrations are excited by motors (vacuum pumps, blowers), so the interference tends to be correlated with the line frequency.

Check with

a) oscilloscope on line trigger

b) hand to feel vibrations

This type of pickup only occurs between conductors at different potentials, so it can be reduced by shielding the relevant electrode.

a) additional shield at electrode potential

b) in coaxial detectors, keep outer electrode at 0 V.

3. RF Pickup

All detector electronics are sensitive to RF signals.

The critical frequency range depends on the shaping time. The gain of the system peaks at

$$f \approx \frac{1}{2\pi t}$$

and high-gain systems will be sensitive over a wide range of frequencies around the peaking frequency.

Typical sources

- Radio and TV stations

  AM broadcast stations: 0.5 – 1.7 MHz

  FM broadcast stations: ~ 100 MHz

  TV stations: 50 – 800 MHz

- Induction furnaces (13.6, 27, 40 MHz)
- Accelerators

- sine waves

- Computers (10's to 100's MHz)
- Video Displays (10 – 100 kHz)
- Radar (GHz)
- Internal clock pulses (e.g. digital control, data readout)

- Pulses (or recurring damped oscillations)

Pulsed UHF or microwave emissions can affect low-frequency circuitry by driving it beyond linearity (the bandwidth of the preamplifier can be much greater than of the subsequent shaper).

Diagnostic Techniques

a) Inspect analog output on an oscilloscope.

Check with different trigger levels and deflection times and look for periodic structure on the baseline.

Pickup levels as low as 10% of the noise level can be serious, so careful adjustment of the trigger and judicious squinting of the eye is necessary to see periodic structure superimposed on the random noise.

Again, an “old fashioned” analog oscilloscope is best.

b) Inspect output with a spectrum analyzer

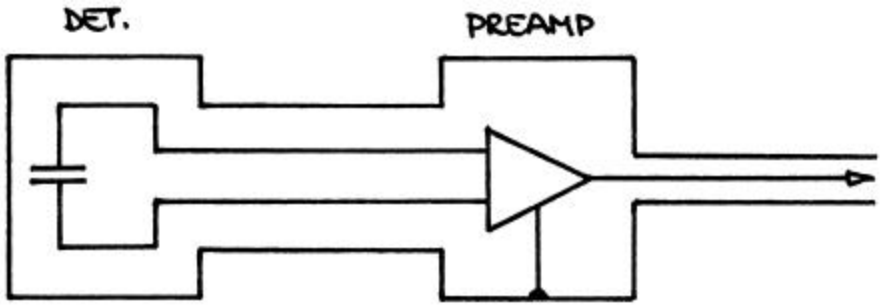
This is a very sensitive technique.

Indeed, for some it may be too sensitive, as it tends to show signals that are so small that they are irrelevant.

Ascertain quantitatively what levels of interfering signals vs. frequency are important.

Remedial Techniques

a) Shielding



contiguous shield

Conducting shield attenuates an incident electromagnetic wave because of

a) reflection of the incident wave

$$E_{0r} = E_0\left(1 - \frac{Z_{shield}}{Z_0}\right)$$

where

$$Z_0 = \sqrt{\frac{m}{e}} = 377\,\Omega$$

is the impedance of free space.

The impedance of the conductor is very low, so most of the incident wave is reflected.

b) attenuation of the absorbed wave

The absorbed wave gives rise to a local current, whose field counteracts the primary excitation.

The net current decreases exponentially as the wave penetrates deeper into the medium

$$i(x) = i_0 e^{-x/d}$$

where i_0 is the current at the surface of the conductor and

$$d = \frac{1}{2 \cdot 10^{-4}} \left[\sqrt{\mathrm{cm} \cdot \mathrm{s}^{-1}} \right] \sqrt{\frac{r}{m_r f}}$$

is the penetration depth or "skin depth". m_r and r are the permeability and resistivity of the conductor and f is the frequency of the incident wave.

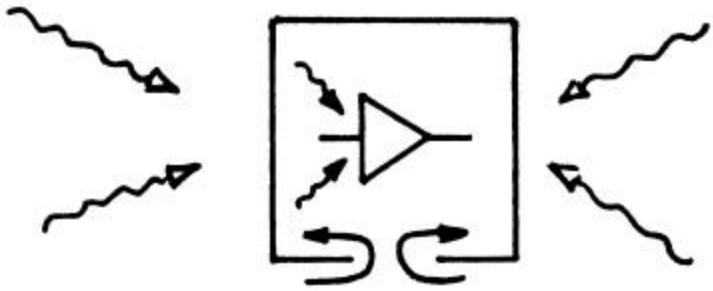
In aluminum, r = 2.8 μΩ·cm and m_r = 1, so
at f = 1 MHz the skin depth d = 84 μm ≈100 μm.

The skin depth decreases with the square root of

- increasing frequency
- decreasing resistivity (increasing conductivity)

If the shield is sufficiently thick, the skin effect isolates the inner surface of a shielding enclosure from the outer surface.

However, this isolation only obtains if no openings in the shield allow the current to flow from the outside to the inside.



External fields can penetrate if openings > l/1000 (diameter of holes, length of slots).

To maintain the integrity of the shield,

- covers must fit tightly with good conductivity at the seams (beware of anodized aluminum!),
- all input and output lines (signal and DC supplies) must have good shield connections,
- shield coverage of coax or other cables >90% and
- connectors must maintain the integrity of the shield connection.

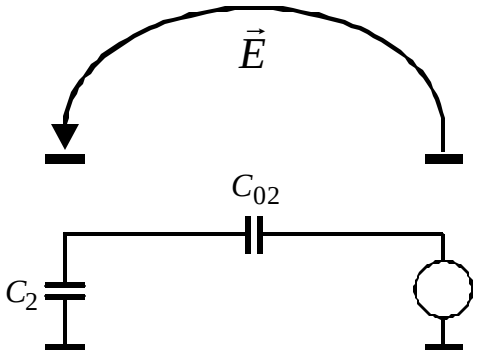
b) “Field Line Pinning”

Full shielding is not always practical, nor is it always necessary.

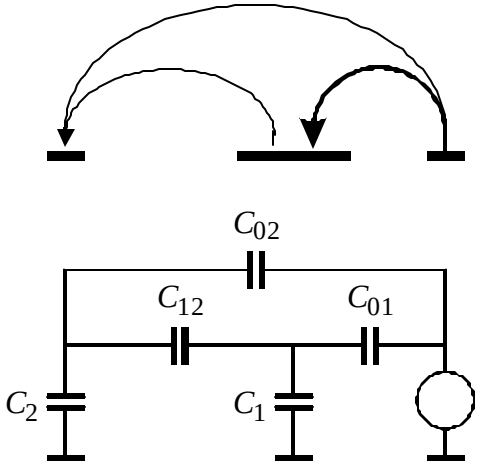
Rather than preventing interference currents from entering the detector system, it is often more practical to reduce the coupling of the interference to the critical nodes.

Consider a conductor carrying an undesired signal current, with a corresponding signal voltage.

Capacitive coupling will transfer interference to another circuit node.



If an intermediate conductor is introduced with a large capacitance to the interference source and to ground compared to the critical node, it will “capture” the field lines and effectively “shield” the critical node.



“Field line pinning” is the operative mechanism of “Faraday shields”.

"Self-Shielding" Structures

The magnitude of capacitive coupling depends on the dielectric constant of the intermediate medium.

Ensemble of electrodes: $\varepsilon_r = 1$ in volume between set 2 to set 1 and $\varepsilon_r > 1$ between sets 2 and 3.

The capacitance between electrode sets 2 and 3 is ε_r times larger than between sets 1 and 2.

Example: Si, $\varepsilon_r = 11.9$

- 92.2% of the field lines originating from electrode set 2 terminate on set 3,
  i.e. are confined to the Si bulk
  7.8% terminate on set 1.

  For comparison, with $\varepsilon_r = 1$, 50% of the field lines originating from electrode set 2 terminate on set 1.

- high dielectric constant reduces coupling of electrode sets 2 and 3 to external sources.

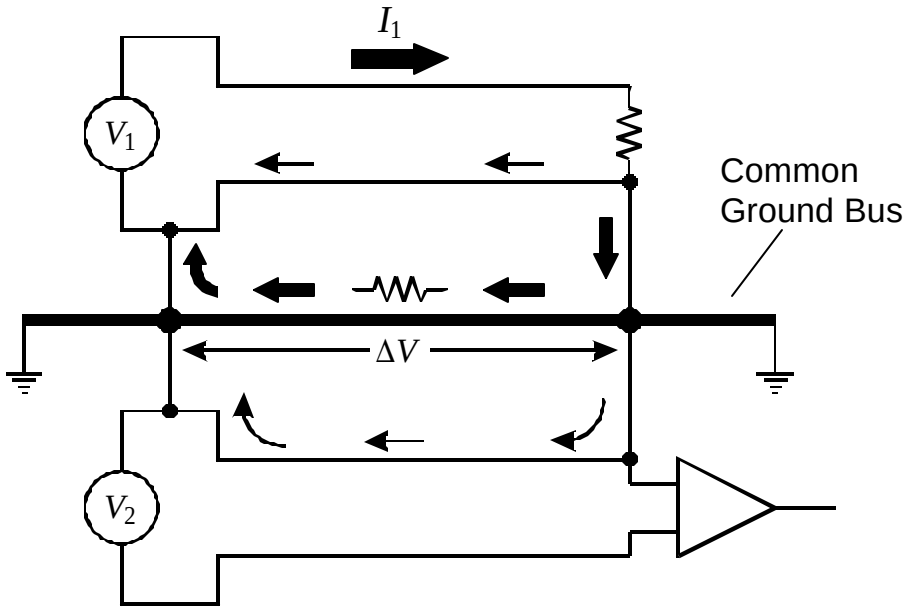
If the interference source is represented by electrode set 1 and sets 2 and 3 represent a detector

- Si detector is 6.5 times less sensitive to capacitive pickup then a detector with $\varepsilon_r = 1$
  (e.g. a gas chamber with the same geometry)

4. Shared Current Paths ("ground loops")

Although capacitive or inductive coupling cannot be ignored, the most prevalent mechanism of undesired signal transfer is the existence of shared signal paths.

Mechanism:



A large alternating current I_1 is coupled into the common ground bus.

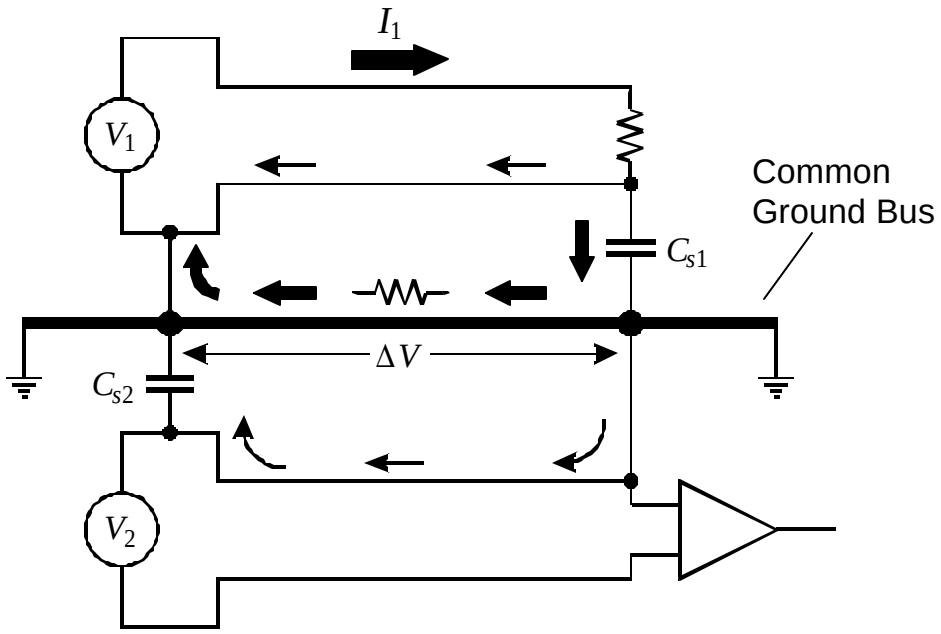
> Although the circuit associated with generator V_1 has a dedicated current return, the current seeks the past of least resistance, which is the massive ground bus.

The lower circuit is a sensitive signal transmission path. Following the common lore, it is connected to ground at both the source and receiver.

The large current flowing through the ground bus causes a voltage drop ΔV, which is superimposed on the low-level signal loop associated with V_2 and appears as an additional signal component.

Cross-coupling has *nothing to do with grounding per se*, but is due to the common return path.
However, the common ground caused the problem by establishing the shared path.

In systems that respond to transients (i.e. time-varying signals) rather than DC signals, secondary loops can be closed by capacitance. A DC path is not necessary.



The loops in this figure are the same as shown before, but the loops are closed by the capacitances C_{s1} and C_{s2}. Frequently, these capacitances are not formed explicitly by capacitors, but are the stray capacitance formed by a power supply to ground, a detector to its support structure (as represented by C_{s2}), etc. For AC signals the inductance of the common current path can increase the impedance substantially beyond the DC resistance, especially at high frequencies.

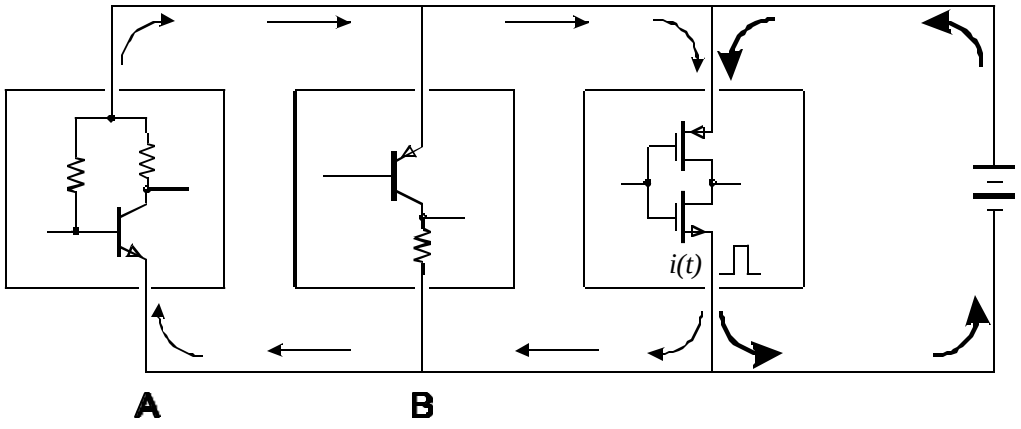
This mode of interference occurs whenever spurious voltages are introduced into the signal path and superimpose on the desired signal.

Interference does not cross-couple by voltage alone, but also via current injection.

Current spikes originating in logic circuitry, for example, propagate through the bussing system as in a transmission line.

Individual connection points will absorb some fraction of the current signal, depending on the relative impedance of the node.

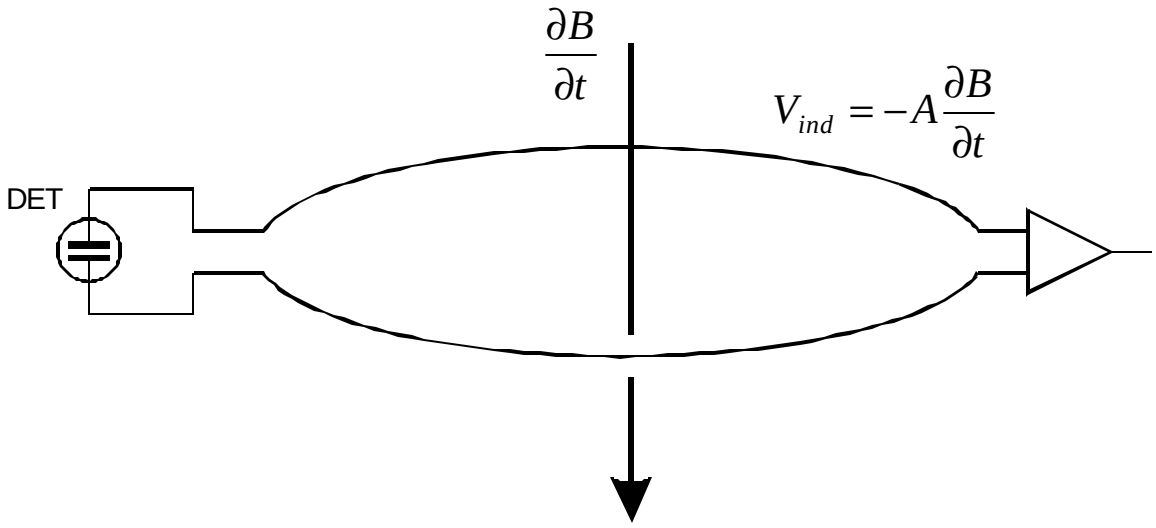
Current spike originates in switching circuit and propagates primarily towards power supply



Current also flows into low impedance node A (common base stage), which closes the secondary loop,

but not into high impedance node B

Another mechanism, beside common conduction paths, is induction:



Clearly, the area A enclosed by any loops should be minimized.

Accomplished by routing signal line and return as a closely spaced pair.

Better yet is a twisted pair, where the voltages induced in successive twists cancel.

Problems occur when alternating detector electrodes are read out at opposite ends – often done because of mechanical constraints.

Remedial Techniques

1. Reduce impedance of the common path

   - Copper Braid Syndrome

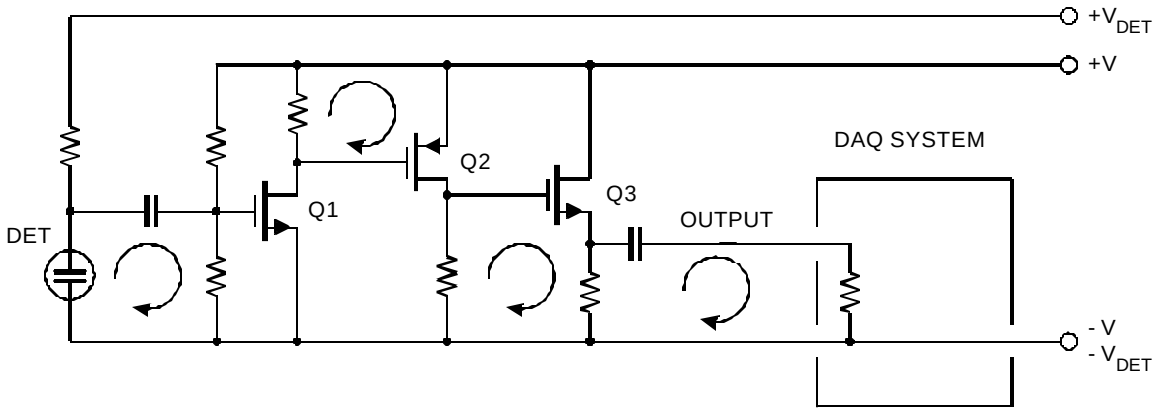
   Colloquially called “improving ground”.

   (sometimes fortuitously introduces an out-of-phase component of the original interference, leading to cancellation)

   Rather haphazard, poorly controlled → continual surprises

2. Avoid Grounds

   Circuits rely on current return paths, not a ground connection!



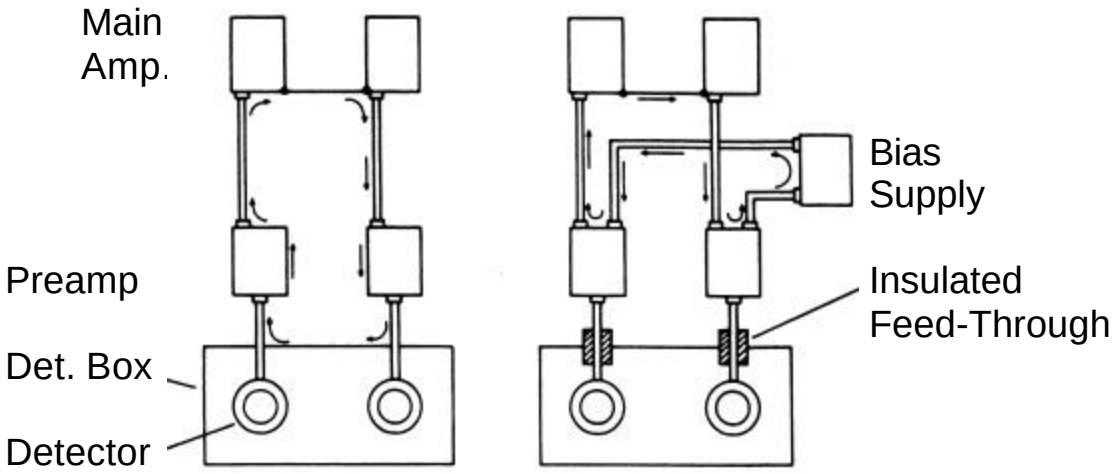
In transferring from stage to stage the signal current flows through local return loops.

1. At the input the detector signal is applied between the gate and source of Q1
2. At the output of Q1 the signal is developed across the load resistor in the drain of Q1 and applied between the gate and source of Q2.
3. The output of Q2 is developed across the load resistor in its drain and applied across the gate and source resistor and load.

Note that – disregarding the input voltage divider that biases Q1 – varying either $+V$ or $-V$ does not affect the local signals.

Breaking parasitic signal paths

Example:



The configuration at the left has a loop that includes the most sensitive part of the system – the detector and preamplifier input.

By introducing insulated feed-throughs, the input loop is broken.

Note that a new loop is shown, introduced by the common detector bias supply. This loop is restricted to the output circuit of the preamplifier, where the signal has been amplified, so it is less sensitive to interference.

- Note that the problem is not caused by loops *per se*, i.e. enclosed areas, but by the multiple connections that provide entry paths for interference.

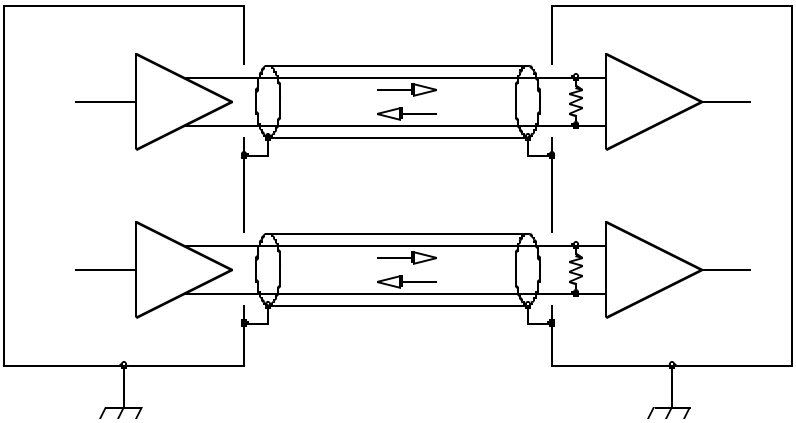
- Although not shown in the schematic illustrations above, both the “detector box” (e.g. a scattering chamber) and the main amplifiers (e.g. in a NIM bin or VME crate) are connected to potential interference sources, so currents can flow through parts of the input signal path.

Breaking shared signal paths, cont'd

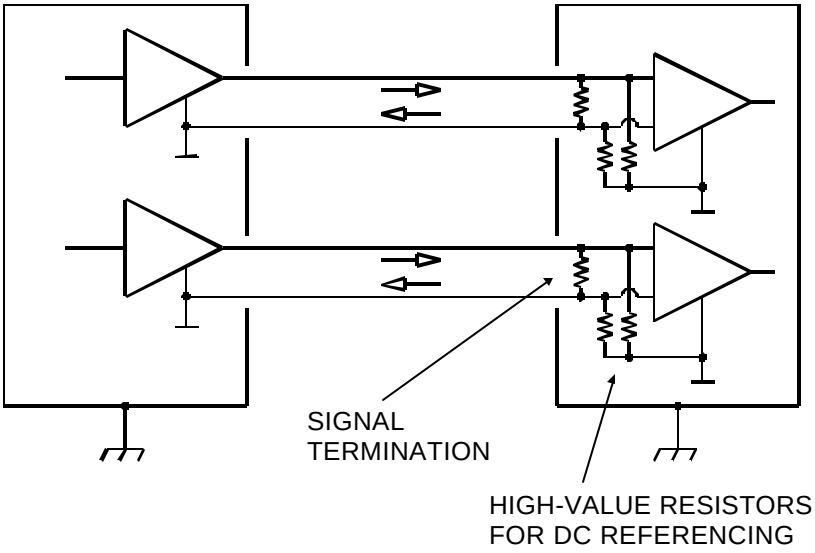
1. Differential Receivers

Besides providing common mode noise rejection, differential receivers also allow "ground free" connections.

Ideal configuration using differential drivers and receivers

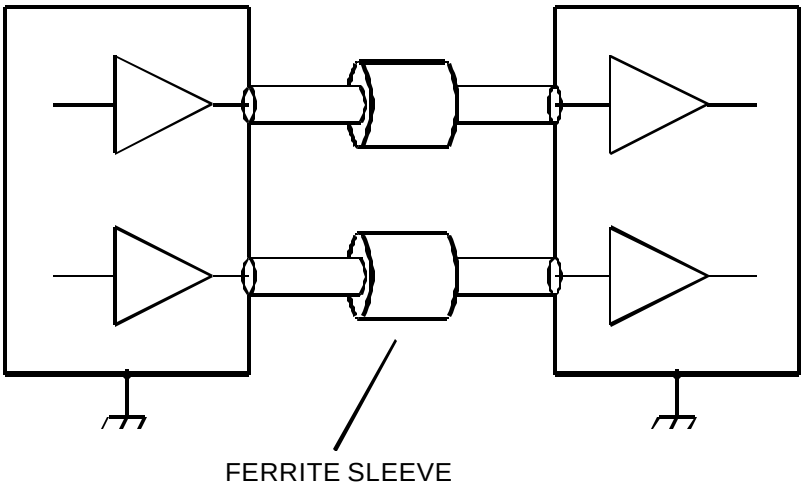


Technique also usable with single-ended drivers



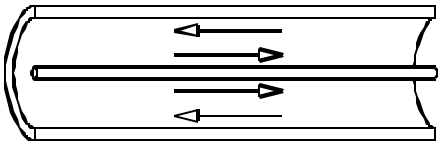
2. Insert high impedances

Ferrite sleeves block common mode currents.



Signal current in coax line flows on

- outer surface of inner conductor
- inner surface of shield.



Net field at *outer surface* of shield is zero.

$\Rightarrow$ Ferrite sleeve does not affect signal transmission.

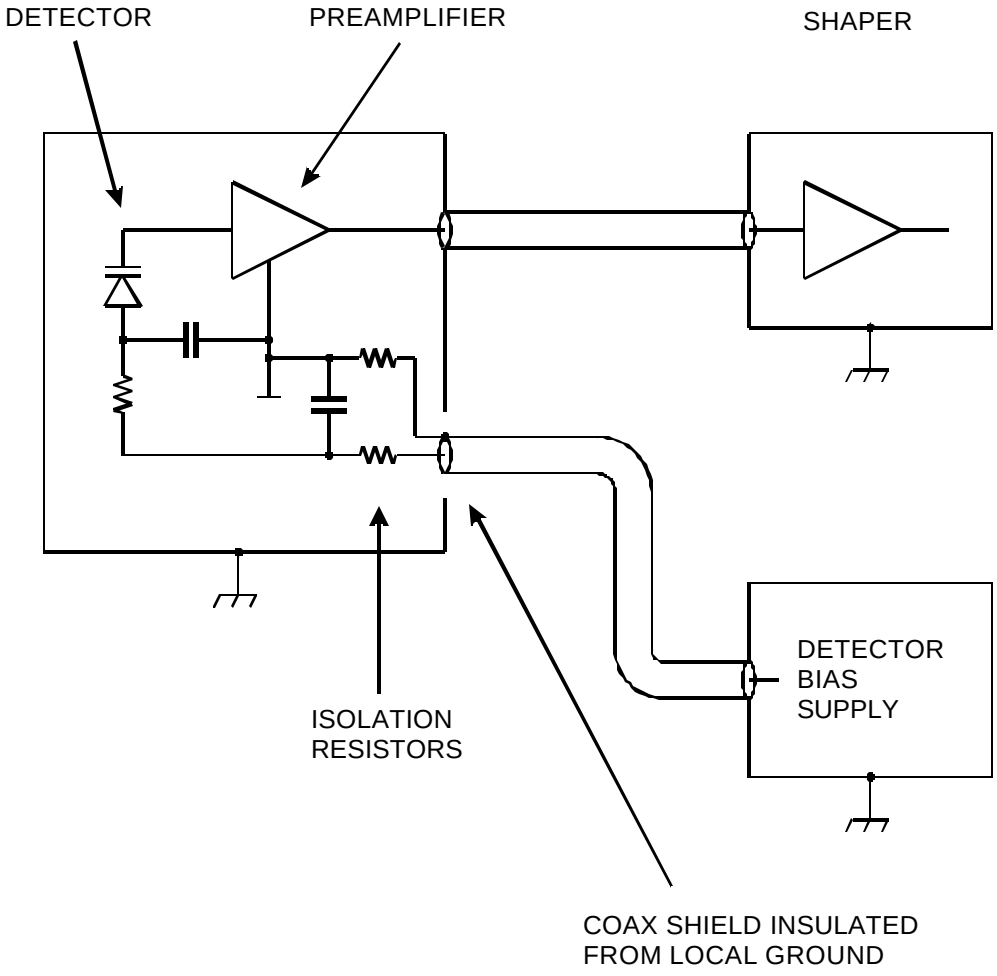
Common mode currents in the coax line
(current flow in same direction on inner and outer conductor) or
current components flowing only on the outside surface of the shield
("ground loops")
will couple to the ferrite and be suppressed.

Ferrite material must be selected to present high impedance at relevant frequencies.

Technique can also be applied to twisted-pair ribbon cables.

Series resistors isolate parasitic ground connections.

Example: detector bias voltage connection



Isolation resistors can also be mounted in an external box that is looped into the bias cable. Either use an insulated box or be sure to isolate the shells of the input and output connectors from another.

A simple check for noise introduced through the detector bias connection is to use a battery.

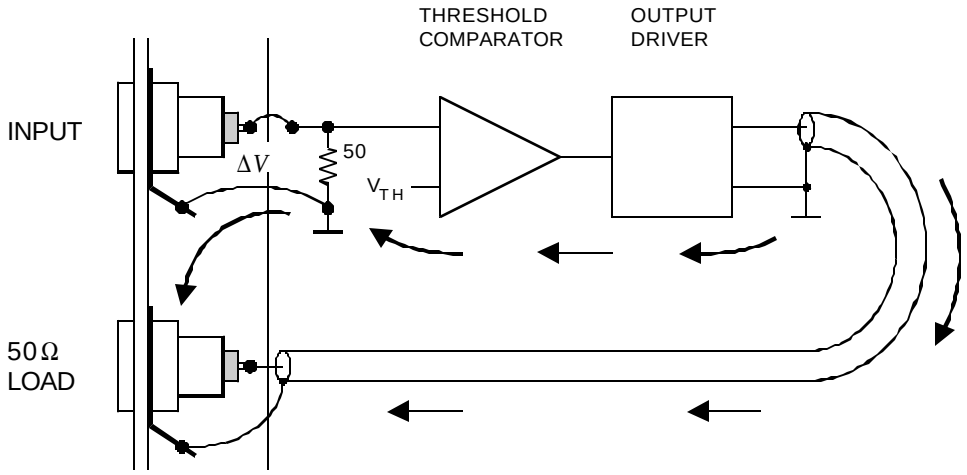
“Ground loops” are often formed by the third wire in the AC power connection. Avoid voltage differences in the “ground” connection by connecting all power cords associated with low-level circuitry into the same outlet strip.

3. Direct the current flow away from sensitive nodes

A timing discriminator was built on a PC board and mounted in a NIM module with multiple channels.

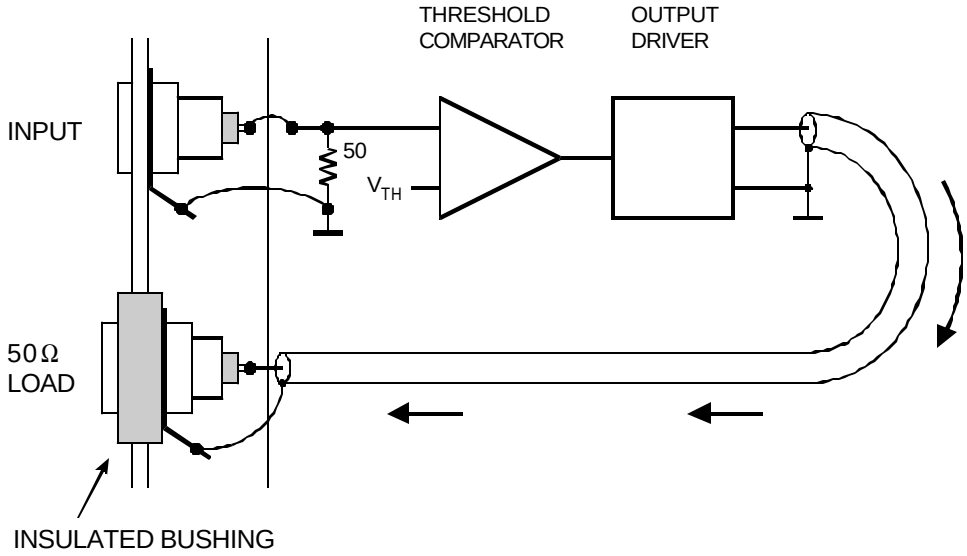
All inputs and outputs were mounted on the front panel. The outputs drove about 20 mA into 50 Ω cables.

Whenever an output fired, the unit broke into oscillation.



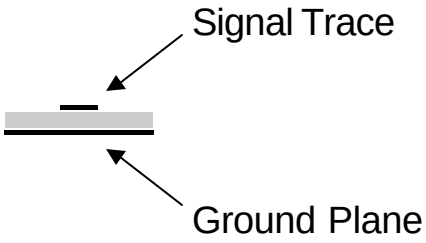
A portion of the output current flowed through the input ground connection. The voltage drop ΔV was sufficient to fire the comparator.

Breaking the loop by insulating the output connector from the front panel fixed the problem.

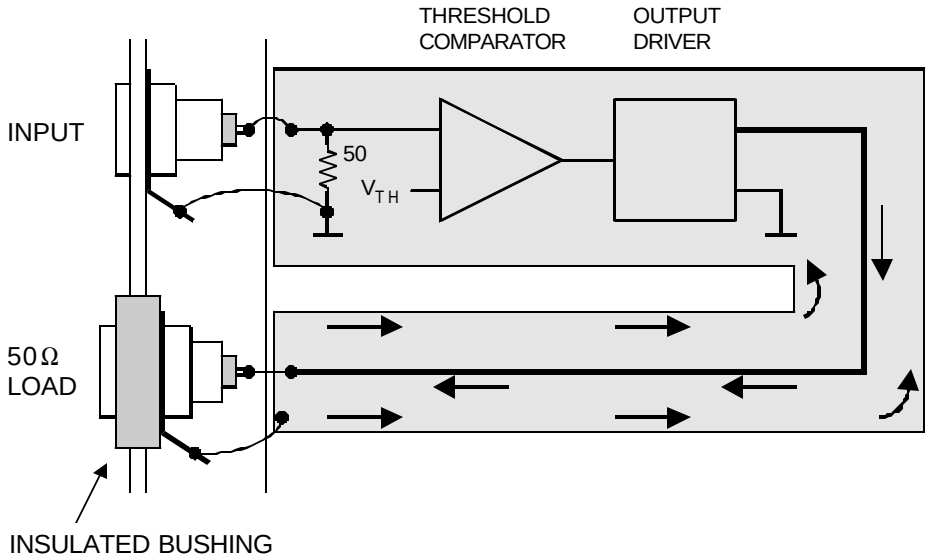


Often it is convenient to replace the coax cable at the output by a strip line integrated on the PC board.

Strip Line:

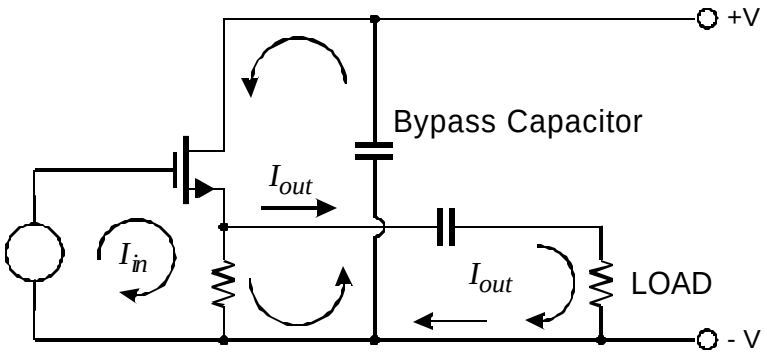


Current paths can be controlled by patterning the ground plane.



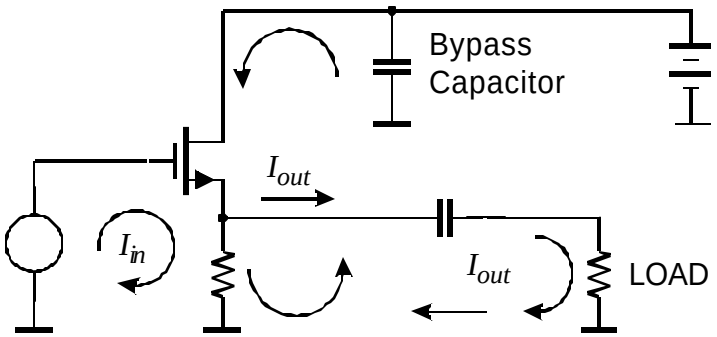
Ground returns are also critical at the circuit level.

Although the desired signal currents circulate as shown on p. 16, there are additional currents flowing in the circuit.



In addition to the signal currents I_{in} and I_{out}, the drain current is also changing with the signal and must return to the source. Since the return through the power supply can be remote and circuitous, a well-defined AC return path is provided by the bypass capacitor.

Since the input and output signal voltages are usually referenced to the negative supply rail, circuits commonly configure it as a common large area bus, the “ground”, and all nodes are referenced to it.



Since the “ground” is a large area conducting surface – often a chassis or a ground plane – with a “low” impedance, it is considered to be an equipotential surface.

The assumption that “ground” is an equipotential surface is not always justified.

At high frequencies current flows only in a thin surface layer (“skin effect”).

The skin depth in aluminum is ~100 μm at 1 MHz. A pulse with a 3 ns rise-time will have substantial Fourier components beyond 100 MHz, where the skin depth is 10 μm.

Even large area conductors can have substantial resistance!

Example: a strip of aluminum, 1 cm wide and 5 cm long has a resistance of ~20 mΩ at 100 MHz (single surface, typical Al alloy)

100 mA → 2 mV voltage drop,

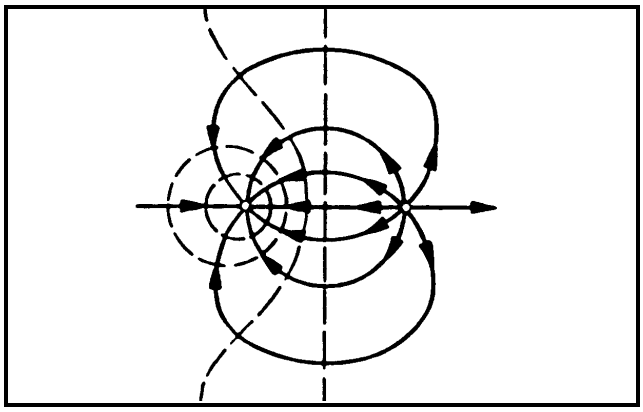
which can be much larger than the signal.

The resistance is determined by the ratio of length to width, i.e. a strip 1 mm wide and 5 mm long will show the same behavior.

Inductance can increase impedances much beyond this value!

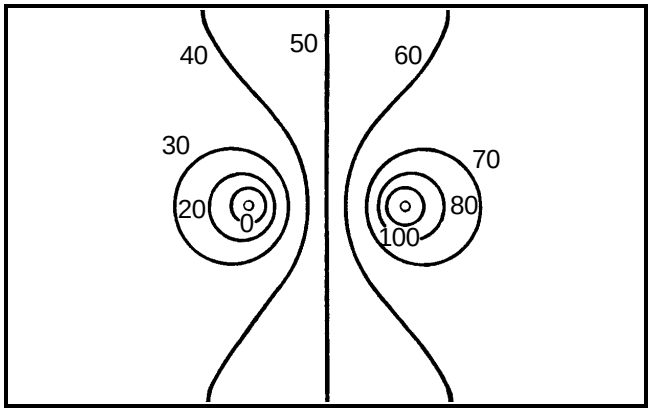
Consider a current loop closed by two connections to a ground plane.

Current distribution around the two connection points:

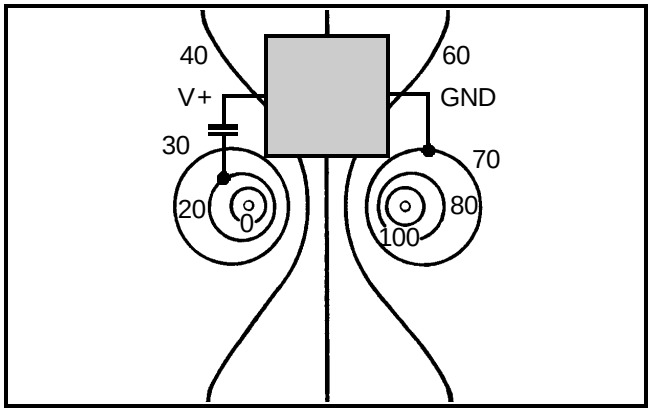


The dashed lines indicate equipotential contours.

Assume a total drop of 100 mV. The resulting potential distribution is

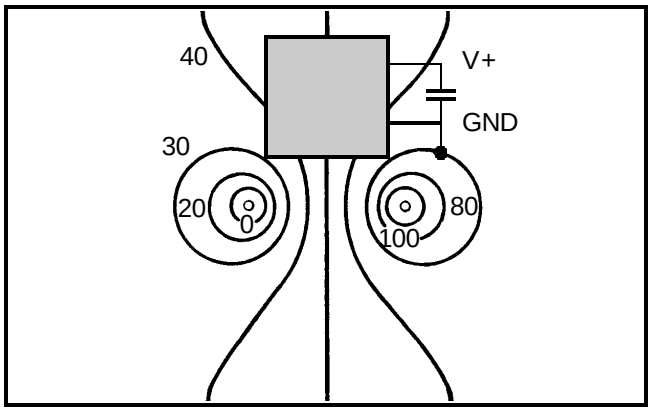


Mounting a circuit block (an IC, for example) with ground and bypass connections as shown below



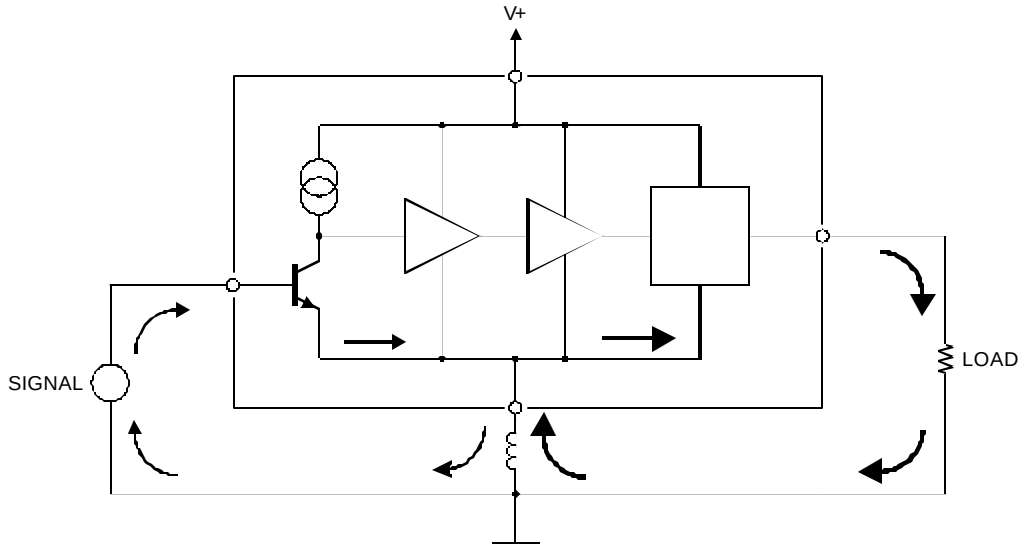
introduces a 50 mV voltage drop in the “ground” path.

Direct connection of the bypass capacitor between the V+ and GND pads avoids pickup of the voltage drop on the ground plane.



“Ground” Connections in Multi-Stage Circuits

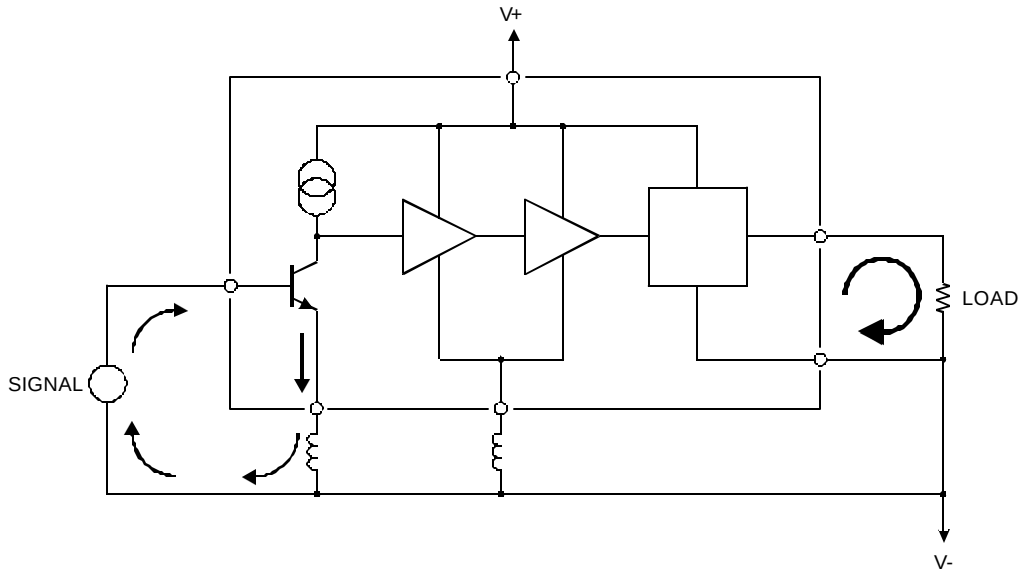
IC comprising a preamplifier, gain stages and an output driver:



The output current is typically orders of magnitude greater than the input current (due to amplifier gain, load impedance).

Combining all ground returns in one bond pad creates a shared impedance (inductance of bond wire). This also illustrates the use of a popular technique – the “star” ground – and its pitfalls.

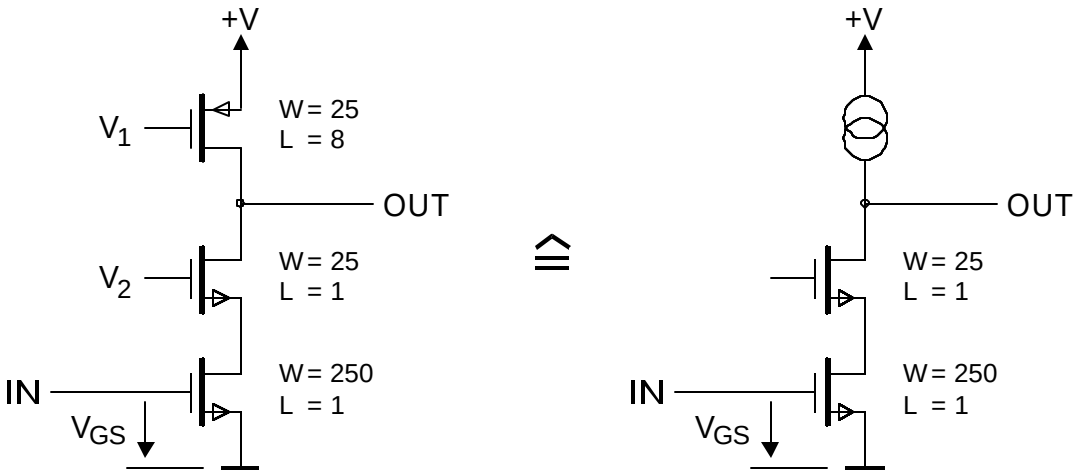
Separating the “ground” connections by current return paths routes currents away from the common impedance and constrains the extent of the output loop, which tends to carry the highest current.



The Folded Cascode

The folded cascode is frequently used in preamplifiers optimized for low power.

Standard cascode with representative transistor sizes:



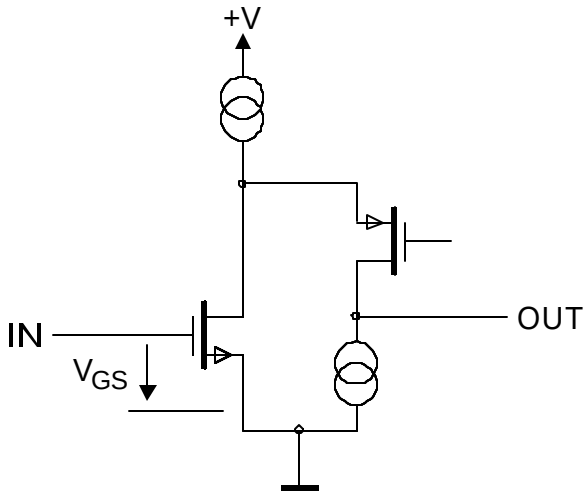
The cascode combines two transistors to obtain

- the high transconductance and low noise of a wide transistor
- the high output resistance (increased by local feedback) and small output capacitance of a narrow transistor
- reduced capacitance between output and input

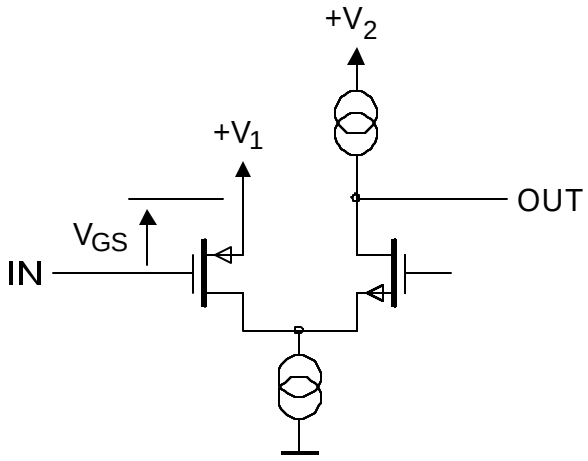
Since the input transistor determines the noise level, its current requirement tends to dominate.

In a conventional cascode the current required for the input transistor must flow through the whole chain.

The folded cascode allows the (second) cascode transistor to operate at a lower current and, as a result, higher output resistance.

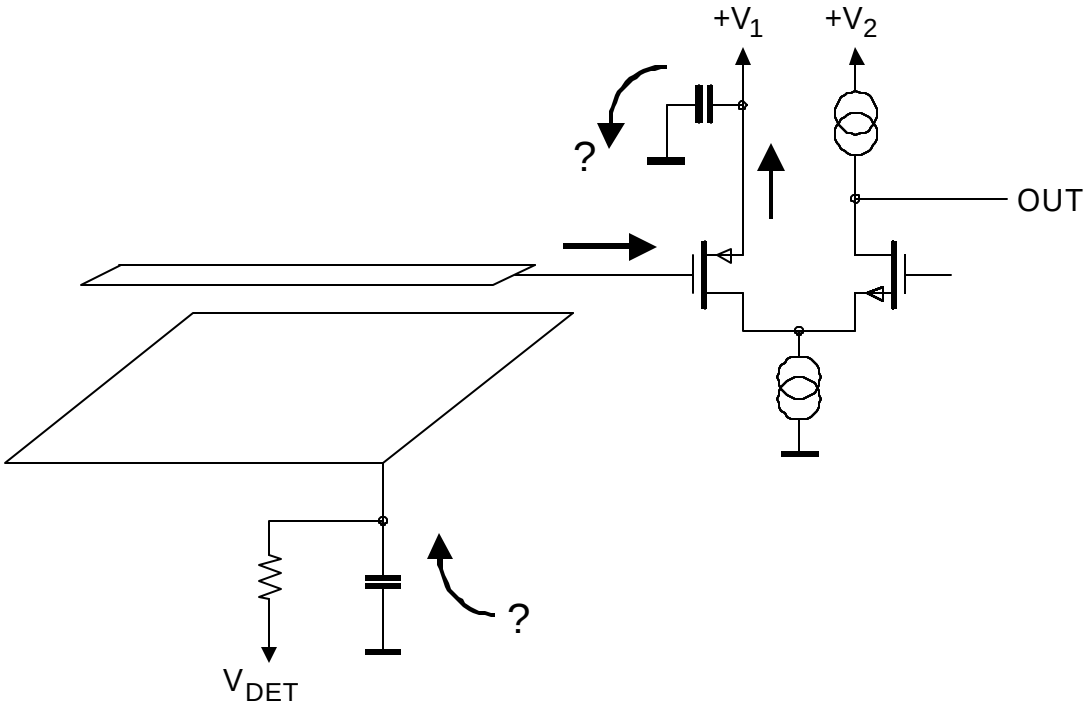


Since PMOS transistors tend to have lower "1/ f " noise than NMOS devices, the following adaptation is often used:



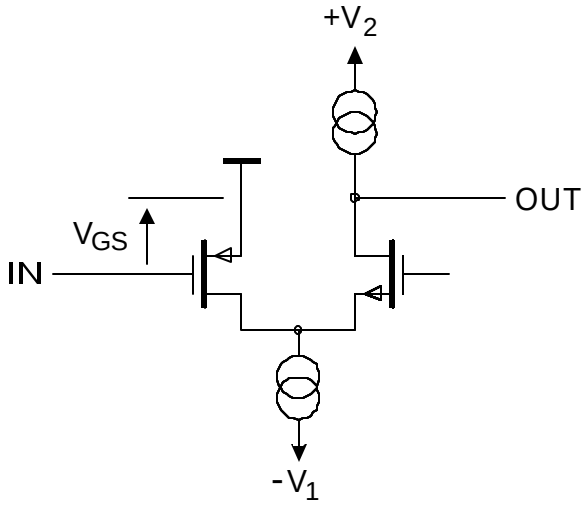
The problem with this configuration is that the supply V_1 becomes part of the input signal path. Unless the V_1 supply bus is very carefully configured and kept free of other signals, interference will be coupled into the input.

Consider the circuit connected to a strip detector:

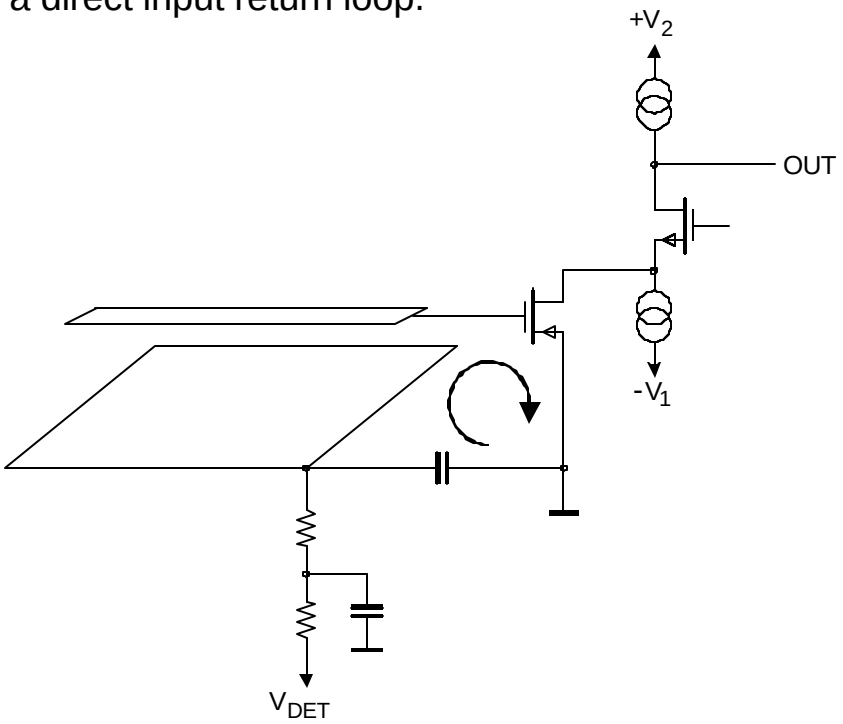


Unless the connection points of the bypass capacitors from the FET source and the detector backplane are chosen carefully, interference will be introduced into the input signal loop.

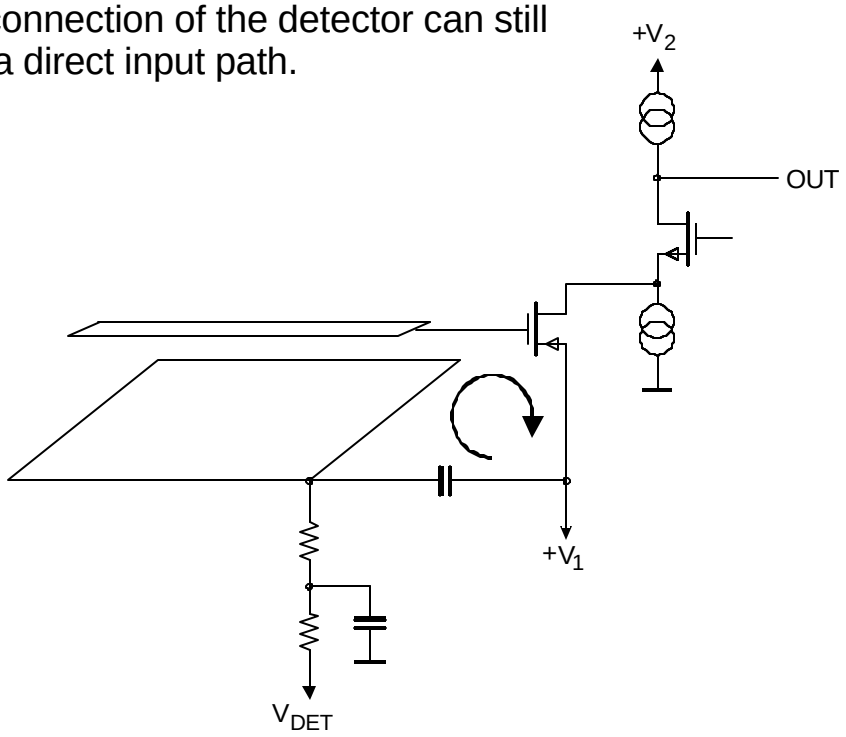
It is much better to “ground” the FET source to a local signal reference and use a negative second supply.



Connected to a strip detector (and redrawn slightly), this configuration provides a direct input return loop.



For some (mythical?) reason positive supplies are more popular. Proper connection of the detector can still provide a direct input path.



In most implementations supply lines are more susceptible to pickup, so the $+V_1$ line must be properly filtered to prevent current injection.

Other Considerations

1. Choice of Shaper

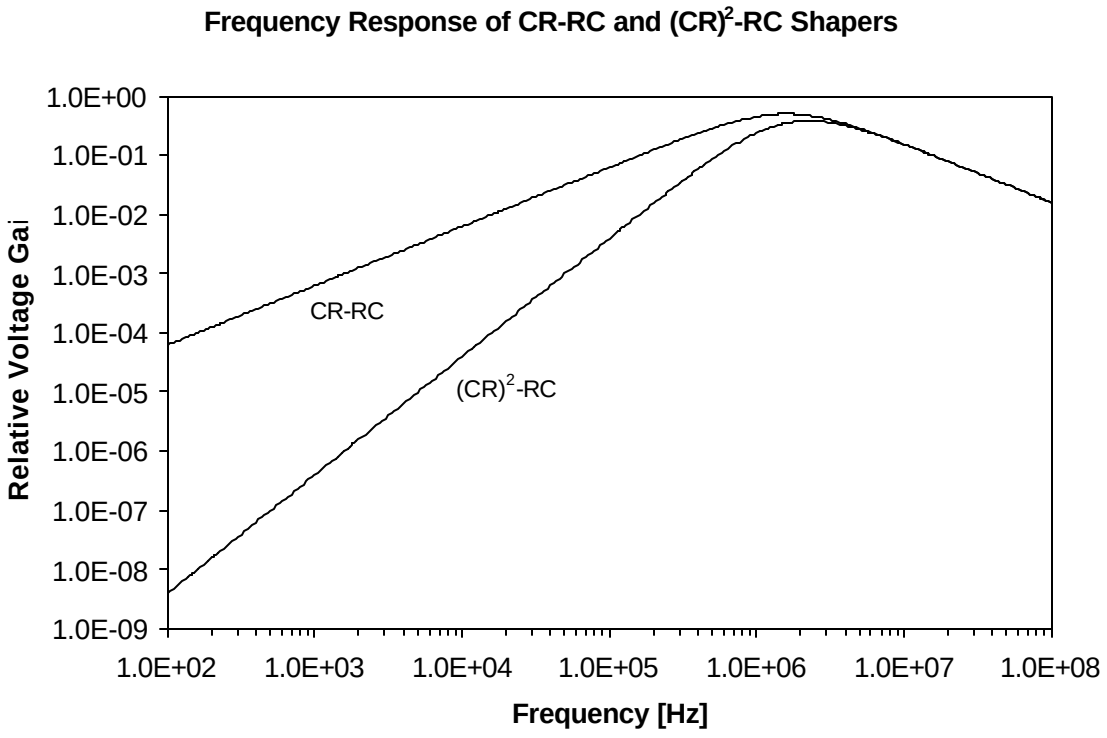
Although a bipolar shaper has slightly inferior noise performance than a unipolar shaper, it may provide better results in the presence of significant low-frequency noise.

Minimum Noise

CR-RC $\quad Q_{n,opt} = 1.355\sqrt{i_n v_n C}$

$(CR)^2$-RC $\quad Q_{n,opt} = 1.406\sqrt{i_n v_n C}$

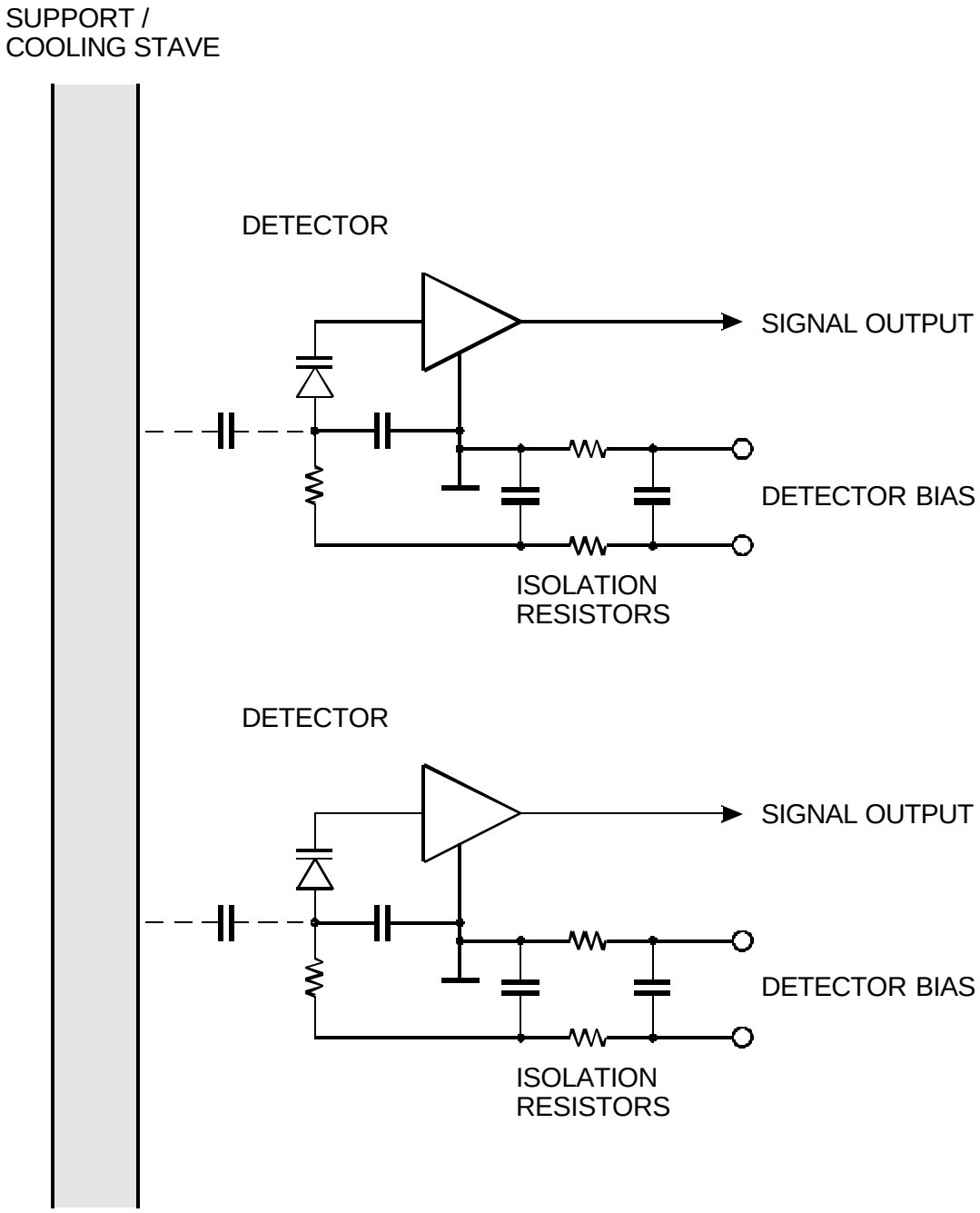
In the frequency domain the additional CR-stage (low-pass filter) provides substantial attenuation of low-frequency interference.



2. Coupling through support structures

Noise currents on the cooling or support structures can couple to the detector input node.

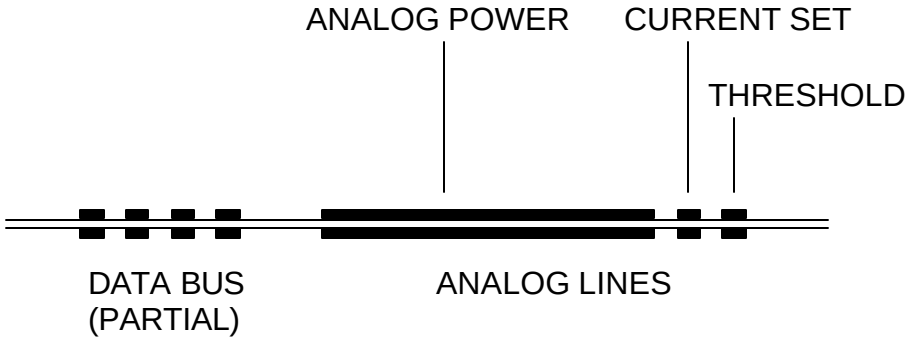
- keep mounting capacitance small, control spurious signals on mount



3. “Self-Shielding” Cables

In mixed analog-digital systems, radiation from cables, especially digital signal cables, is a concern.

By utilizing broadside-coupled differential lines with a thin intermediate dielectric, the field is confined to the region between the conductors. The extent of the fringing field beyond the conductor edge is about equal to the thickness of the dielectric.



Example:

Conductors:	50 μm thick, 150 μm wide
Dielectric:	50 μm thick
Gap between pairs:	150 μm

Cross-coupling between adjacent pairs:
< 2% at 60 MHz for 1 m length

Power connections are made substantially wider (1 – 5 mm), forming a low impedance transmission line with high distributed capacitance.

The geometry shown above is for short runs in the inner region of a tracker, where reduction of material is crucial. Dimensions can be scaled proportionally to achieve lower resistance and signal dispersion in longer cable runs at larger radii.

8. Summary of Considerations in Detector Electronics

1. Maximize the signal

   Maximizing the signal also implies reducing the capacitance at the electronic input node. Although we want to measure charge, the primary electric signal is either voltage or current, both of which increase with decreasing capacitance.

2. Choose the input transistor to match the application.

   At long shaping times FETs (JFETs or MOSFETs) are best. At short shaping times, bipolar transistors tend to prevail.

3. Select the appropriate shaper and shaping time

   In general, short shaping times will require higher power dissipation for a given noise level than long times.

   The shaper can be optimized with respect to either current or voltage noise (important in systems subject to radiation damage)

   The choice of shaping function and time can significantly affect the sensitivity to external pickup.

4. Position-sensitive detectors can be implemented using either interpolation techniques or direct readout. Interpolating systems reduce the number of electronic channels but require more complex and sophisticated electronics. Direct readout allows the greatest simplicity per channel, but requires many channels, often at high density (good match for monolithically integrated circuits).

5. Segmentation improves both rate capability and noise (low capacitance). It also increases radiation resistance.

6. Timing systems depend on slope-to-noise ratio, so they need to optimize both rise-time and capacitance.

   Relatively long rise-times can still provide good timing resolution (<< rise-time), if the signal-to-noise ratio is high.

   Variations in signal transit times and pulse shape can degrade time resolution significantly.

7. Electronic noise in practical systems can be predicted and understood *quantitatively*.

8. From the outset, systems must consider sensitivity to spurious signals and robustness against self-oscillation.

   Poor system configurations can render the best low-noise front-end useless, but proper design can yield “laboratory” performance in large-scale systems.

9. Although making detectors “work” in an experiment has relied extensively on tinkering and “cut-and-try”, understanding the critical elements that determine detector performance makes it much easier to navigate the maze of a large system.

   It is more efficient to avoid problems than to fix them.

   A little understanding can go a long way.

Appendix 1: Noise Spectral Densities

Spectral Density of Thermal Noise

Two approaches can be used to derive the spectral distribution of thermal noise.

1. The thermal velocity distribution of the charge carriers is used to calculate the time dependence of the induced current, which is then transformed into the frequency domain.

2. Application of Planck's theory of black body radiation.

The first approach clearly shows the underlying physics, whereas the second "hides" the physics by applying a general result of statistical mechnics. However, the first requires some advanced concepts that go well beyond the standard curriculum, so the "black body" approach will be used.

In Planck's theory of black body radiation the energy per mode

$$\bar{E} = \frac{h\nu}{e^{h\nu/kT} - 1}$$

and the spectral density of the radiated power

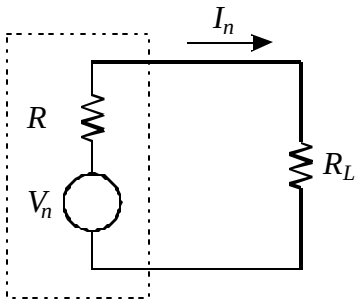
$$\frac{dP}{d\nu} = \frac{h\nu}{e^{h\nu/kT} - 1}$$

i.e. this is the power that can be extracted in equilibrium. At low frequencies $h\nu \ll kT$

$$\frac{dP}{d\nu} \approx \frac{h\nu}{\left(1 + \frac{h\nu}{kT}\right) - 1} = kT \ ,$$

so at low frequencies the spectral density is independent of frequency and for a total bandwidth B the noise power that can be transferred to an external device $P_n = kTB$.

To apply this result to the noise of a resistor, consider a resistor R whose thermal noise gives rise to a noise voltage V_n. To determine the power transferred to an external device consider the circuit



The power dissipated in the load resistor R_L

$$\frac{V_{nL}^2}{R_L} = I_n^2 R_L = \frac{V_n^2 R_L}{(R+R_L)^2}$$

The maximum power transfer occurs when the load resistance equals the source resistance $R_T = R$, so

$$V_{nL}^2 = \frac{V_n^2}{4}$$

Since the power transferred to R_L is kTB

$$\frac{V_{nL}^2}{R} = \frac{V_n^2}{4R} = kTB$$

$$P_n = \frac{V_n^2}{R} = 4kTB$$

and the spectral density of the noise power

$$\frac{dP_n}{dn} = 4kT$$

Spectral Density of Shot Noise

If an excess electron is injected into a device, it forms a current pulse of duration t. In a thermionic diode t is the transit time from cathode to anode (see IX.2), for example. In a semiconductor diode t is the recombination time (see IX-2). If these times are short with respect to the periods of interest $t \ll 1/f$, the current pulse can be represented by a d pulse. The Fourier transform of a delta pulse yields a "white" spectrum, i.e. the amplitude distribution in frequency is uniform

$$\frac{dI_{npk}}{df} = 2q_e$$

Within an infinitesimally narrow frequency band the individual spectral components are pure sinusoids, so their rms value

$$i_n \equiv \frac{dI_n}{df} = \frac{2q_e}{\sqrt{2}} = \sqrt{2}q_e$$

If N electrons are emitted at the same average rate, but at different times, they will have the same spectral distribution, but the coefficients will differ in phase. For example, for two currents i_p and i_q with a relative phase j the total rms current

$$\left\langle i^2 \right\rangle = \left(i_p + i_q e^{ij}\right)\left(i_p + i_q e^{-ij}\right) = i_p^2 + i_q^2 + 2i_p i_q \cos j$$

For a random phase the third term averages to zero

$$\left\langle i^2 \right\rangle = i_p^2 + i_q^2 \ ,$$

so if N electrons are randomly emitted per unit time, the individual spectral components simply add in quadrature

$$i_n^2 = 2Nq_e^2$$

The average current

$$I = Nq_e \ ,$$

so the spectral noise density

$$i_n^2 \equiv \frac{dI_n^2}{df} = 2q_e I$$

“Noiseless” Resistances

a) Dynamic Resistance

In many instances a resistance is formed by the slope of a device’s current-voltage characteristic, rather than by a static ensemble of electrons agitated by thermal energy.

Example: forward-biased semiconductor diode

Diode current vs. voltage

$$I = I_0(e^{q_e V / kT} - 1)$$

The differential resistance

$$r_d = \frac{dV}{dI} = \frac{kT}{q_e I}$$

i.e. at a given current the diode presents a resistance, e.g. 26 Ω at I = 1 mA and T = 300 K.

Note that two diodes can have different charge carrier concentrations, but will still exhibit the same dynamic resistance at a given current, so the dynamic resistance is not uniquely determined by the number of carriers, as in a resistor.

There is no thermal noise associated with this “dynamic” resistance, although the current flow carries shot noise.

b) Radiation Resistance of an Antenna

Consider a receiving antenna with the normalized power pattern $P_n(q,f)$ pointing at a brightness distribution $B(q,f)$ in the sky. The power per unit bandwidth received by the antenna

$$w = \frac{A_e}{2}\iint B(q,f)P_n(q,f)d\Omega$$

where A_e is the effective aperture, i.e. the "capture area" of the antenna. For a given field strength E, the captured power $W \propto EA_e$.

If the brightness distribution is from a black body radiator and we're measuring in the Rayleigh-Jeans regime,

$$B(q,f) = \frac{2kT}{l^2}$$

and the power received by the antenna

$$w = \frac{kT}{l^2}A_e\Omega_A \ .$$

Ω_A is the beam solid angle of the antenna (measured in rad^2), i.e. the angle through which all the power would flow if the antenna pattern were uniform over its beamwidth.

Since $A_e\Omega_A = l^2$ (see antenna textbooks), the received power

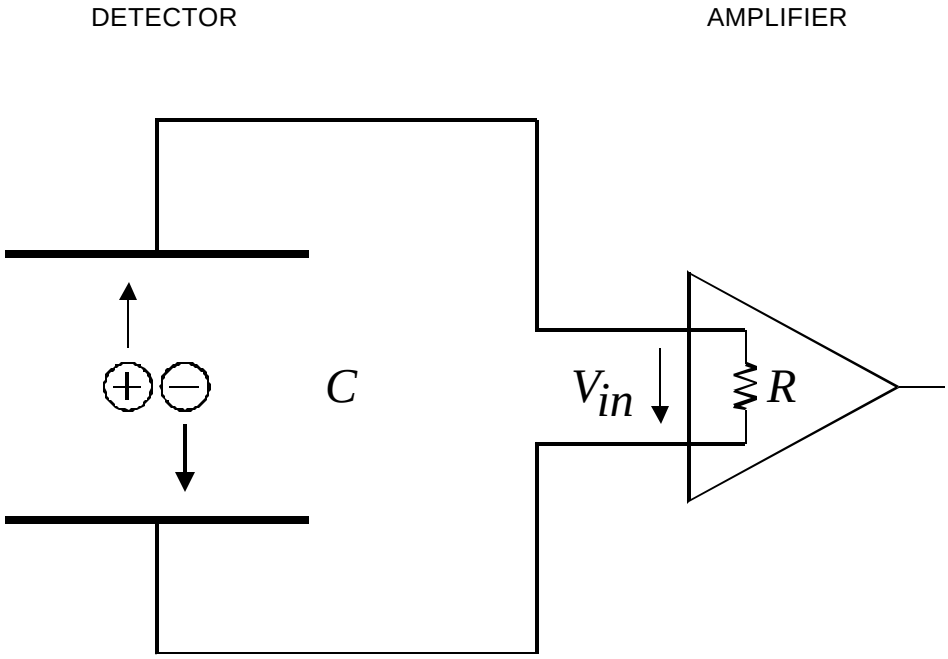
$$w = kT$$

The received power is independent of the radiation resistance, as would be expected for thermal noise.

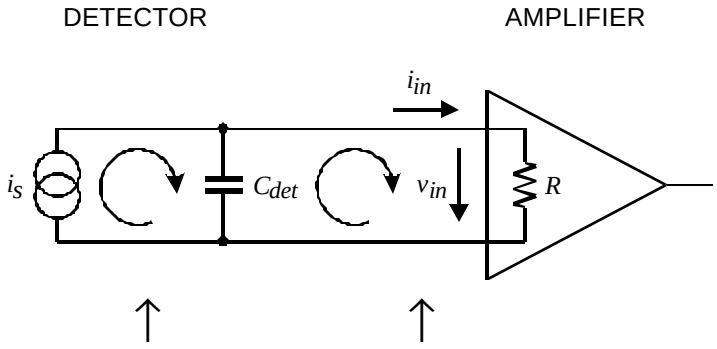
However, it is not determined by the temperature of the antenna, but by the temperature of the sky the antenna pattern is subtending.

For example, for a region dominated by the CMB, the measured power corresponds to a resistor at a temperature of ~3K, although the antenna may be at 300K.

Appendix 2
Signal-to-Noise Ratio vs. Detector Capacitance



Equivalent Circuit



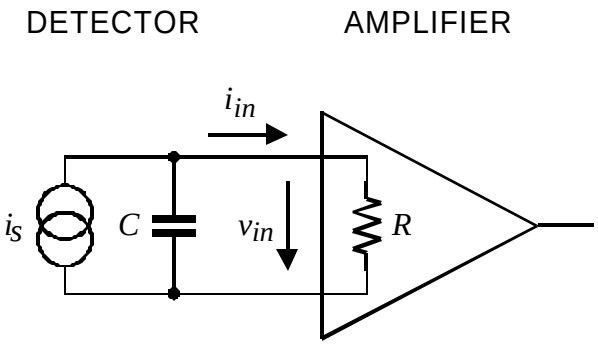
charges moving in detector induce change of charge on detector electrodes

detector capacitance discharges into amplifier

Assume an amplifier with constant noise. Then signal-to-noise ratio (and the equivalent noise charge) depend on the signal magnitude. Pulse shape registered by amplifier depends on the input time constant RC_{det}.

Assume a rectangular detector current pulse of duration T and magnitude I_s.

Equivalent circuit



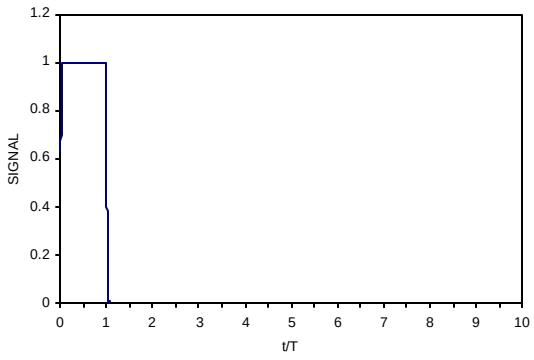
Input current to amplifier

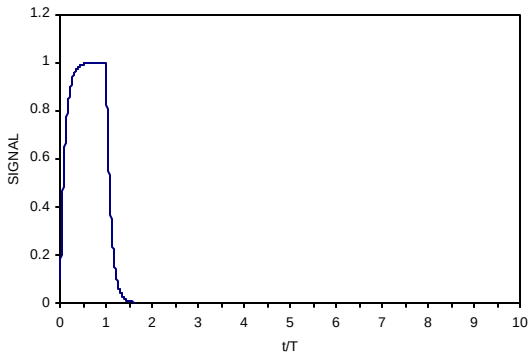
$$0 \le t < T: \qquad i_{in}(t) = I_s\left(1 - e^{-t/RC}\right)$$

$$T \le t \le \infty: \qquad i_{in}(t) = I_s\left(e^{T/RC} - 1\right)\cdot e^{-t/RC}$$

At short time constants $RC << T$ the amplifier pulse approximately follows the detector current pulse.

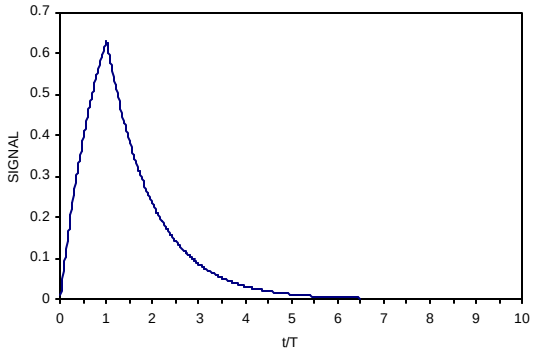
$RC = 0.01\ T$ $RC = 0.1\ T$



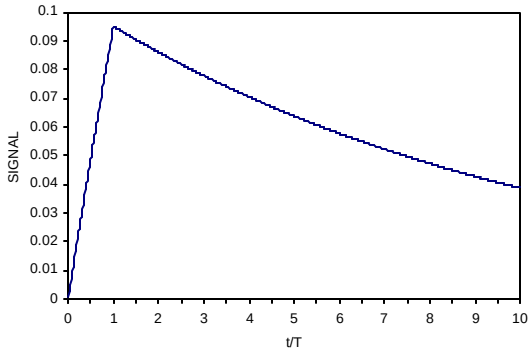


As the input time constant RC increases, the amplifier signal becomes longer and the peak amplitude decreases, although the integral, i.e. the signal charge, remains the same.

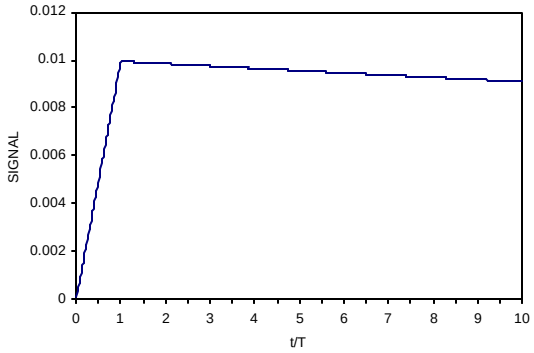
$RC = T$



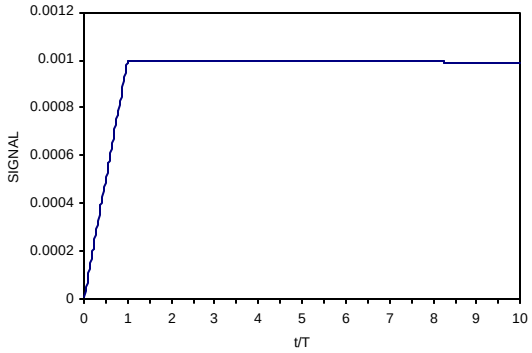
$RC = 10\ T$



$RC = 100\ T$



$RC = 10^3\ T$



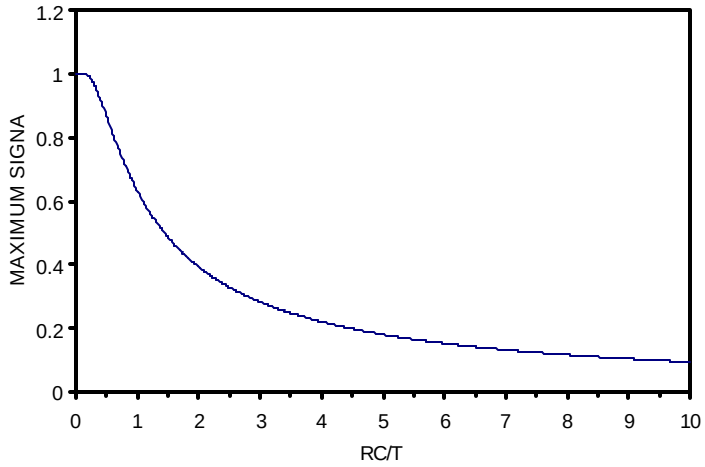
At long time constants the detector signal current is integrated on the detector capacitance and the resulting voltage sensed by the amplifier

$$V_{in} = \frac{Q_{\text{det}}}{C} = \frac{\int i_s dt}{C}$$

Then the peak amplifier signal is inversely proportional to the **total capacitance at the input**, i.e. the sum of

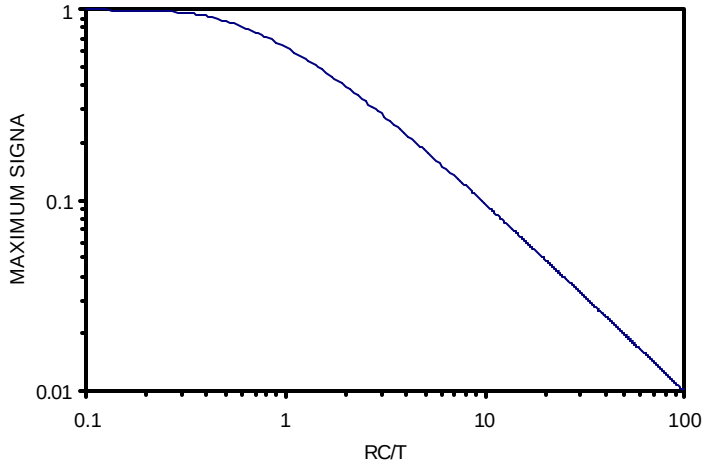
- detector capacitance,
- input capacitance of the amplifier, and
- stray capacitances.

Maximum signal vs. capacitance



At small time constants the amplifier signal approximates the detector current pulse and is independent of capacitance.

At large input time constants ($RC/T > 5$) the maximum signal falls linearly with capacitance.



- For input time constants large compared to the detector pulse duration the signal-to-noise ratio decreases with detector capacitance.

Caution when extrapolating to smaller capacitances:

If $S/N = 1$ at $RC/T = 100$, decreasing the capacitance to 1/10 of its original value ($RC/T = 10$), increases S/N to 10.

However, if initially $RC/T = 1$, the same 10-fold reduction in capacitance (to $RC/T = 0.1$) only yields $S/N = 1.6$.

Appendix 3: Noise in Transistors

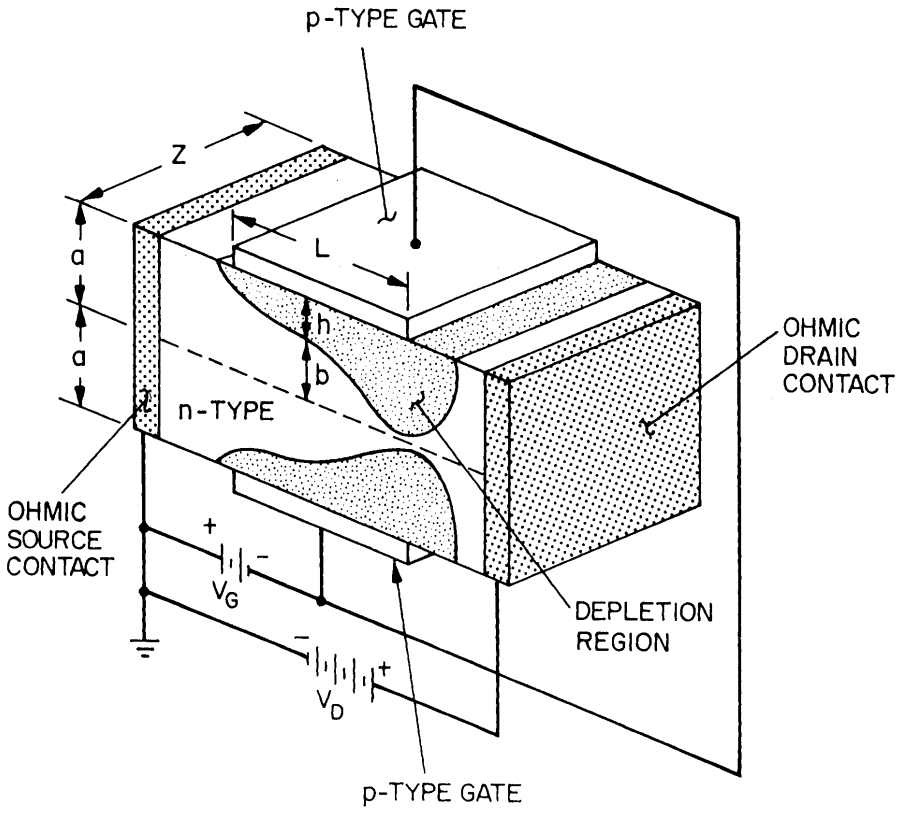
a) Field Effect Transistors

Field Effect Transistors (FETs) utilize a conductive channel whose resistance is controlled by an applied potential.

1. Junction Field Effect Transistor (JFET)

In JFETs a conducting channel is formed of n or p-type semiconductor (GaAs, Ge or Si).
Connections are made to each end of the channel, the Drain and Source.

In the implementation shown below a pair of gate electrodes of opposite doping with respect to the channel are placed at opposite sides of the channel. Applying a reverse bias forms a depletion region that reduces the cross section of the conducting channel.



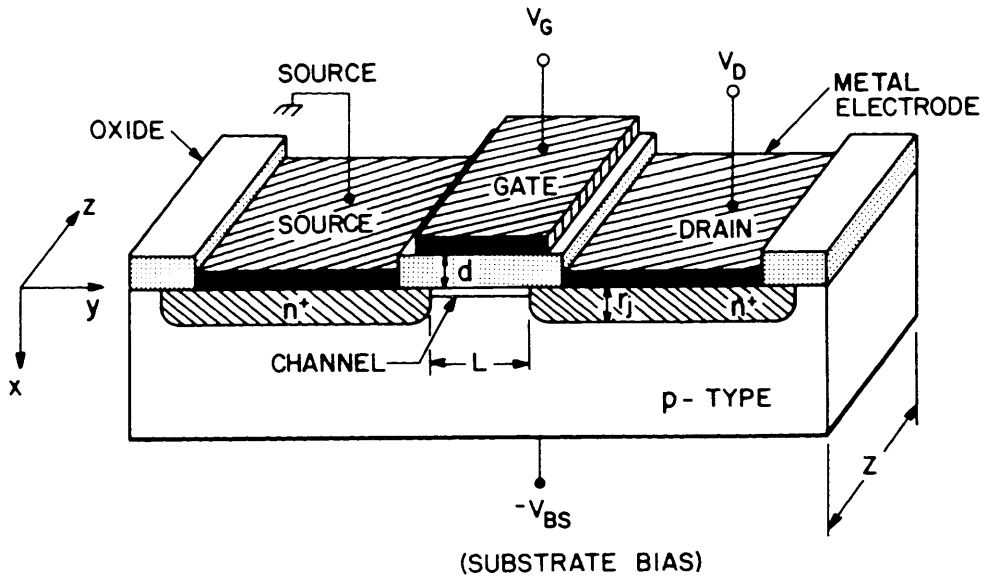
(from Sze)

Changing the magnitude of the reverse bias on the gate modulates the cross section of the channel.

2. Metal Oxide Field Effect Transistors (MOSFETs)

Both JFETs and MOSFETs are conductivity modulated devices, utilizing only one type of charge carrier. Thus they are called unipolar devices, unlike bipolar transistors, for which both electrons and holes are crucial.

Unlike a JFET, where a conducting channel is formed by doping and its geometry modulated by the applied voltages, the MOSFET changes the carrier concentration in the channel.



(from Sze)

The source and drain are n^+ regions in a p-substrate.

The gate is capacitively coupled to the channel region through an insulating layer, typically SiO_2.

Applying a positive voltage to the gate increases the electron concentration at the silicon surface beneath the gate.

- As in a JFET the combination of gate and drain voltages control the conductivity of the channel.

- Both JFETs and MOSFETs are characterized primarily by transconductance, i.e. the change in output current vs. input voltage

a) Noise in Field Effect Transistors

The primary noise sources in field effect transistors are

a) thermal noise in the channel

b) gate current in JFETs

Since the area of the gate is small, this contribution to the noise is very small and usually can be neglected.

Thermal velocity fluctuations of the charge carriers in the channel superimpose a noise current on the output current.

The spectral density of the noise current at the drain is

$$i_{nd}^2 = \frac{N_{C,tot} q_e}{L^2} m_0 \, 4k_B T_e$$

The current fluctuations depend on the number of charge carriers in the channel $N_{C,tot}$ and their thermal velocity, which in turn depends on their temperature T_e and low field mobility m_0. Finally, the induced current scales with 1/L because of Ramo's theorem.

To make practical use of the above expression it is necessary to express it in terms of directly measureable device parameters. Since the transconductance in the saturation region

$$g_m \propto \frac{W}{L} m N_{ch} d$$

one can express the noise current as

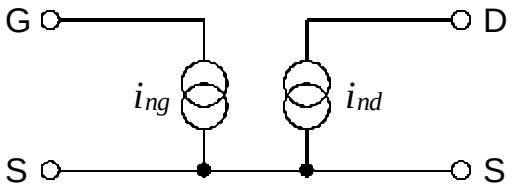
$$i_{nd}^2 = g_n g_m \, 4k_B T_0$$

where T_0= 300K and g_n is a semi-empirical constant that depends on the carrier concentration in the channel and the device geometry.

In a JFET the gate noise current is the shot noise associated with the reverse bias current of the gate-channel diode

$$i_{ng} = 2q_e I_G$$

The noise model of the FET



The gate and drain noise currents are independent of one another.

However, if an impedance Z is connected between the gate and the source, the gate noise current will flow through this impedance and generate a voltage at the gate

$$e_{ng} = Z\ i_{ng}$$

leading to an additional noise current at the output $g_m v_{ng}$, so that the total noise current at the output becomes

$$i_{no}^2 = i_{nd}^2 + (g_m Z i_{ng})^2$$

To allow a direct comparison with the input signal this cumulative noise will be referred back to the input to yield the equivalent input noise voltage

$$e_{ni}^2 = \frac{i_{no}^2}{g_m^2} = \frac{i_{nd}^2}{g_m^2} + Z i_{ng}^2 \equiv e_n^2 + Z i_n^2$$

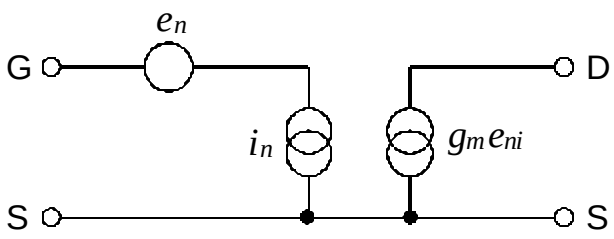
i.e. referred to the input, the drain noise current i_{nd} translates into a noise voltage source

$$e_n^2 = 4k_B T_0 \frac{g_n}{g_m}$$

The noise coefficient g_n is usually given as 2/3, but is typically in the range 0.5 to 1 (exp. data will shown later).

This expression describes the noise of both JFETs and MOSFETs.

In this parameterization the noise model becomes



where e_n and i_n are the input voltage and current noise. As was shown above, these contribute to the input noise voltage e_{ni}, which in turn translates to the output through the transconductance g_m to yield a noise current at the output $g_m e_{ni}$.

The equivalent noise charge

$$Q_n^2 = i_n^2 F_i T + e_n^2 C_i^2 \frac{F_v}{T}$$

For a representative JFET g_m= 0.02, C_i= 10 pF and I_G < 150 pA. If F_i=F_v=1

$$Q_n^2 = 1.9 \cdot 10^9 T + \frac{3.25 \cdot 10^{-3}}{T}$$

As the shaping time T is reduced, the current noise contribution decreases and the voltage noise contribution increases. For T= 1 µs the current contribution is 43 el and the voltage contribution 3250 el, so the current contribution is negligible, except in very low frequency applications.

Optimization of Device Geometry

For a given device technology and normalized operating current I_D/W both the transconductance and the input capacitance are proportional to device width W

$$g_m \propto W \quad \text{and} \quad C_i \propto W$$

so that the ratio

$$\frac{g_m}{C_i} = const$$

Then the signal-to-noise ratio can be written as

$$\left(\frac{S}{N}\right)^2 = \frac{(Q_s/C)^2}{v_n^2} = \frac{Q_s^2}{(C_{\det} + C_i)^2}\frac{g_m}{4k_B T_0 \Delta f}$$

$$\left(\frac{S}{N}\right)^2 = \frac{Q_s^2}{\Delta f}\frac{1}{4k_B T_0}\left(\frac{g_m}{C_i}\right)\frac{1}{C_i\left(1+\frac{C_{\det}}{C_i}\right)^2}$$

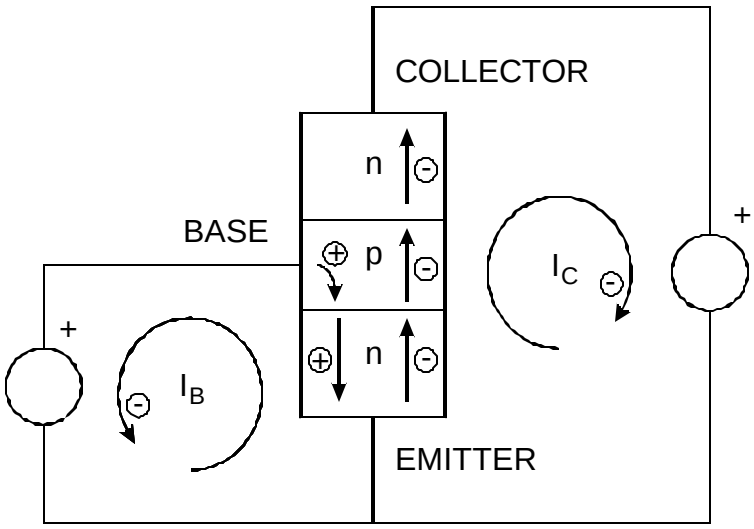
S/N is maximized for $C_i = C_{det}$ (capacitive matching).

$C_i << C_{det}$: The detector capacitance dominates, so the effect of increased transistor capacitance is negligible.
As the device width is increased the transconductance increases and the equivalent noise voltage decreases, so S/N improves.

$C_i > C_{det}$: The equivalent input noise voltage decreases as the device width is increased, but only with $1/\sqrt{W}$, so the increase in capacitance overrides, decreasing S/N.

Bipolar Transistors

Consider the *npn* structure shown below.



The base and emitter form a diode, which is forward biased so that a base current I_B flows.

The base current injects holes into the base-emitter junction.

As in a simple diode, this gives rise to a corresponding electron current through the base-emitter junction.

If the potential applied to the collector is sufficiently positive so that the electrons passing from the emitter to the base are driven towards the collector, an external current I_C will flow in the collector circuit.

The ratio of collector to base current is equal to the ratio of electron to hole currents traversing the base-emitter junction. In an ideal diode

$$\frac{I_C}{I_B} = \frac{I_{nBE}}{I_{pBE}} = \frac{D_n / N_A L_n}{D_p / N_D L_p} = \frac{N_D}{N_A} \frac{D_n L_p}{D_p L_n}$$

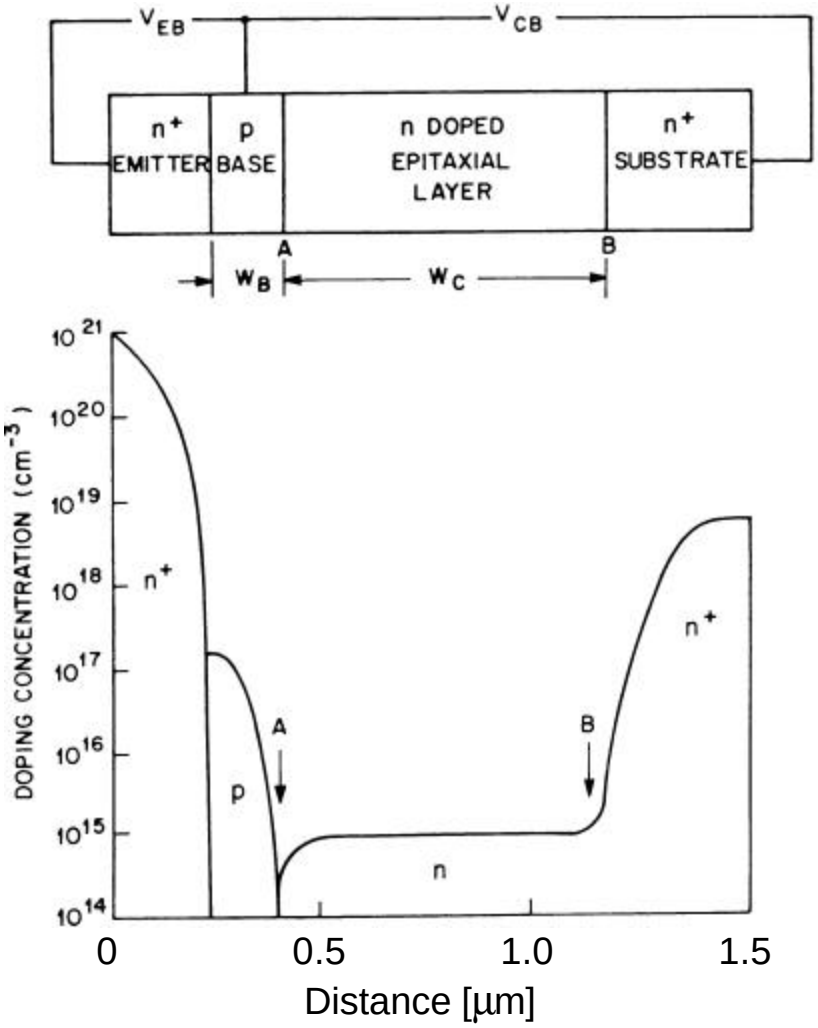
If the ratio of doping concentrations in the emitter and base regions N_D/N_A is sufficiently large, the collector current will be greater than the base current.

- DC current gain

Furthermore, we expect the collector current to saturate when the collector voltage becomes large enough to capture all of the minority carrier electrons injected into the base.

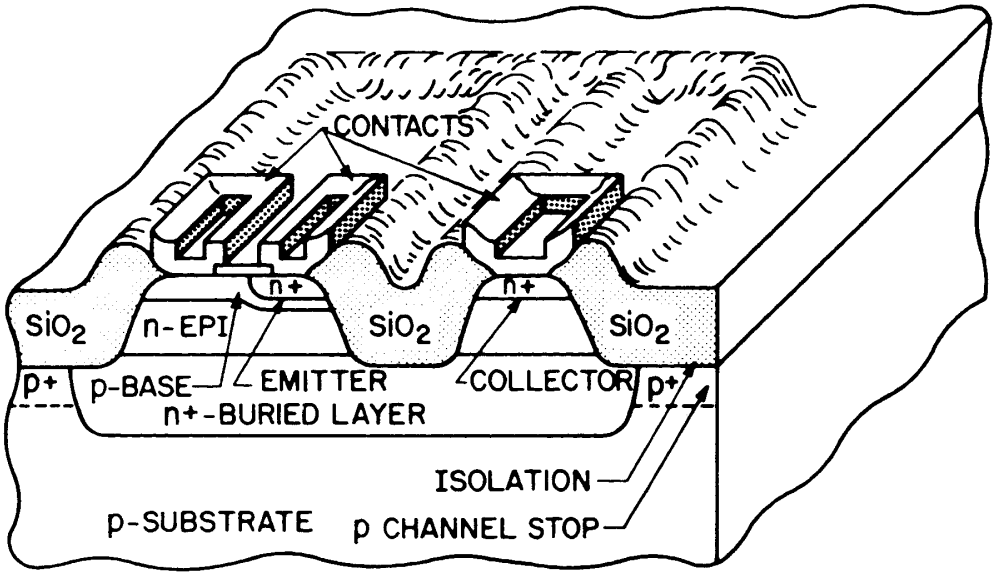
Since the current inside the transistor comprises both electrons and holes, the device is called a bipolar transistor.

Dimensions and doping levels of a modern high-frequency transistor (5 – 10 GHz bandwidth)



(adapted from Sze)

High-speed bipolar transistors are implemented as vertical structures.



(from Sze)

The base width, typically 0.2 μm or less in modern high-speed transistors, is determined by the difference in diffusion depths of the emitter and base regions.

The thin base geometry and high doping levels make the base-emitter junction sensitive to large reverse voltages.

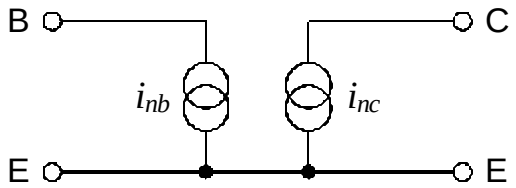
Typically, base-emitter breakdown voltages for high-frequency transistors are but a few volts.

As shown in the preceding figure, the collector region is usually implemented as two regions: one with low doping (denoted “epitaxial layer” in the figure) and the other closest to the collector contact with a high doping level. This structure improves the collector voltage breakdown characteristics.

b) Noise in Bipolar Transistors

In bipolar transistors the shot noise from the base current is important.

The basic noise model is the same as shown before, but the magnitude of the input noise current is much greater, as the base current will be 1 – 100 μA rather than <100 pA.



The base current noise is shot noise associated with the component of the emitter current provided by the base.

$$i_{nb}^2 = 2q_e I_B$$

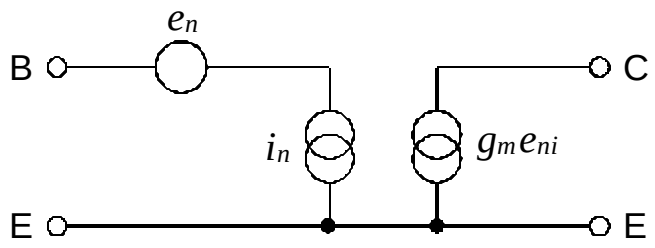
The noise current in the collector is the shot noise originating in the base-emitter junction associated with the collector component of the emitter current.

$$i_{nc}^2 = 2q_e I_C$$

Following the same argument as in the analysis of the FET, the output noise current is equivalent to an equivalent noise voltage

$$e_n^2 = \frac{i_{nc}^2}{g_m^2} = \frac{2q_e I_C}{(q_e I_C / k_B T)^2} = \frac{2(k_B T)^2}{q_e I_C}$$

yielding the noise equivalent circuit



where i_n is the base current shot noise i_{nb}.

The equivalent noise charge

$$Q_n^2 = i_n^2 F_i T + e_n^2 C^2 \frac{F_v}{T} = 2q_e I_B F_i T + \frac{2(k_B T)^2}{q_e I_C} C^2 \frac{F_v}{T}$$

Since $I_B = I_C / b_{DC}$

$$Q_n^2 = 2q_e \frac{I_C}{b_{DC}} F_i T + \frac{2(k_B T)^2}{q_e I_C} C^2 \frac{F_v}{T}$$

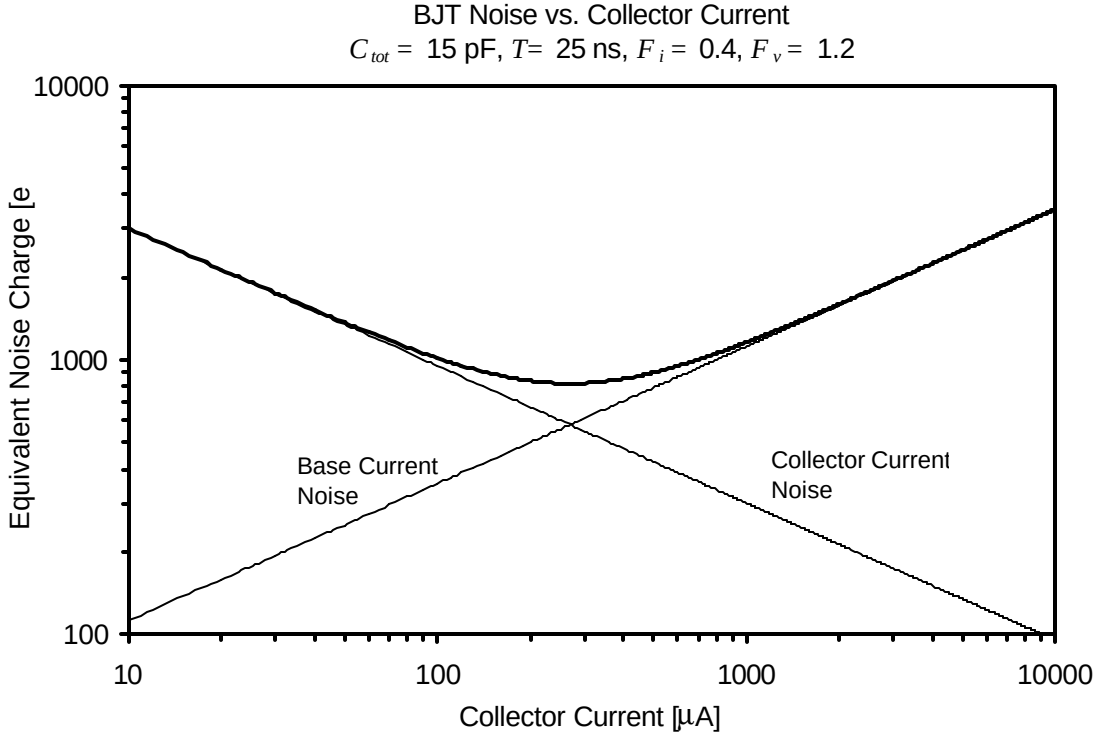
The current noise term increases with I_C, whereas the second (voltage) noise term decreases with I_C.

Thus, the noise attains a minimum

$$Q_{n,\min}^2 = 4k_B T \frac{C}{\sqrt{b_{DC}}} \sqrt{F_i F_v}$$

at a collector current

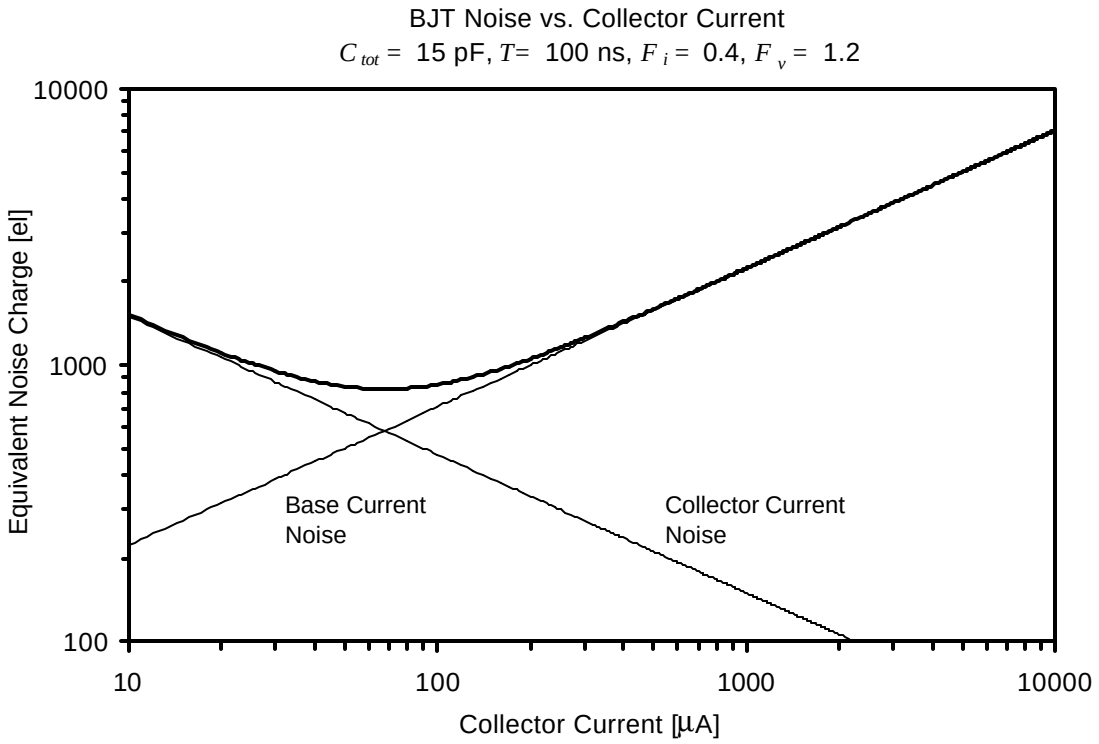
$$I_C = \frac{k_B T}{q_e} C \sqrt{b_{DC}} \sqrt{\frac{F_v}{F_i}} \frac{1}{T} .$$



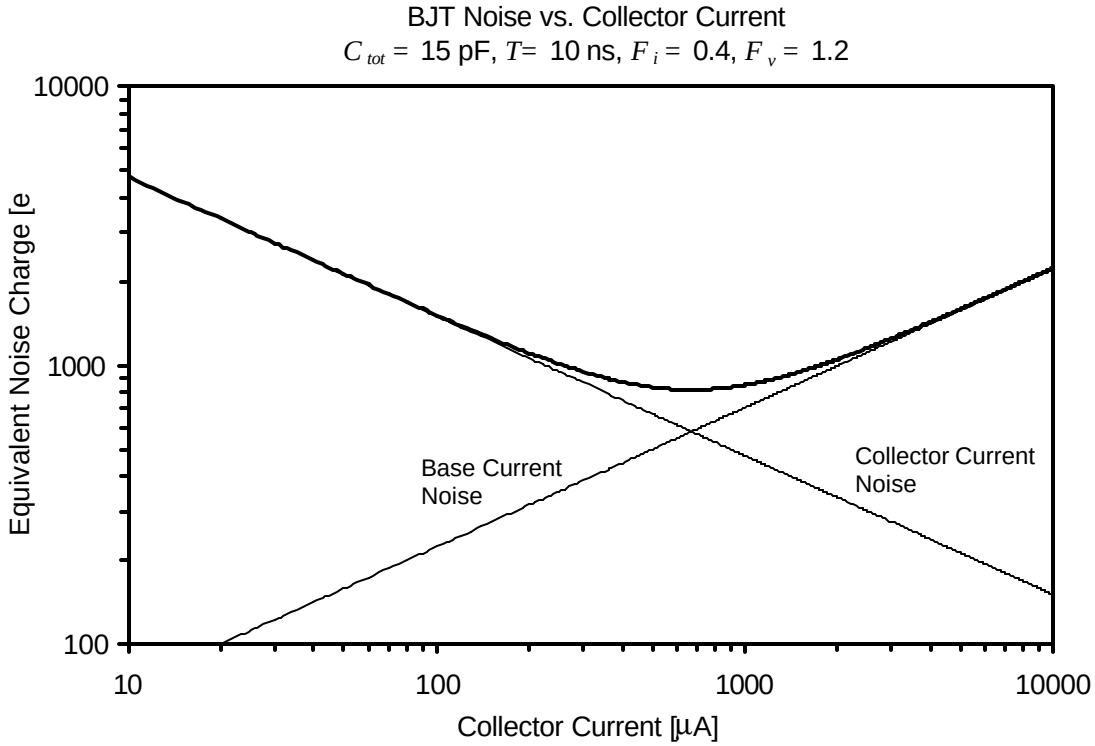
- For a given shaper, the minimum obtainable noise is determined only by the total capacitance at the input and the DC current gain of the transistor, *not by the shaping time*.

- The shaping time only determines the current at which this minimum noise is obtained

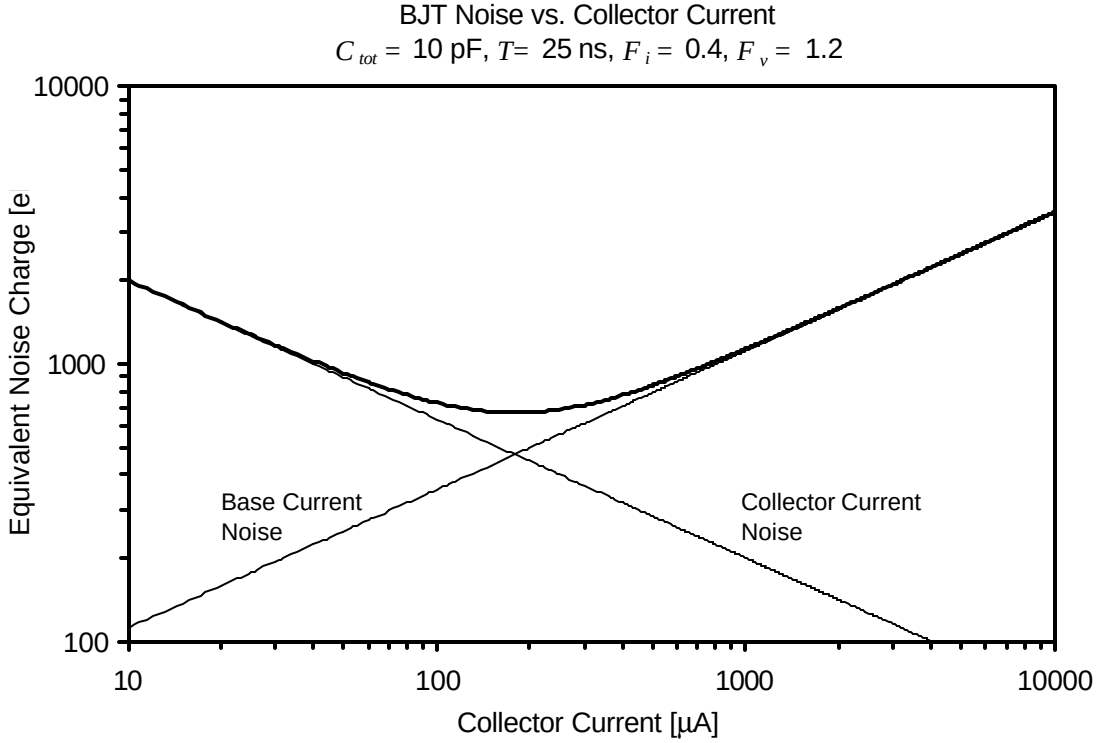
T= 100 ns

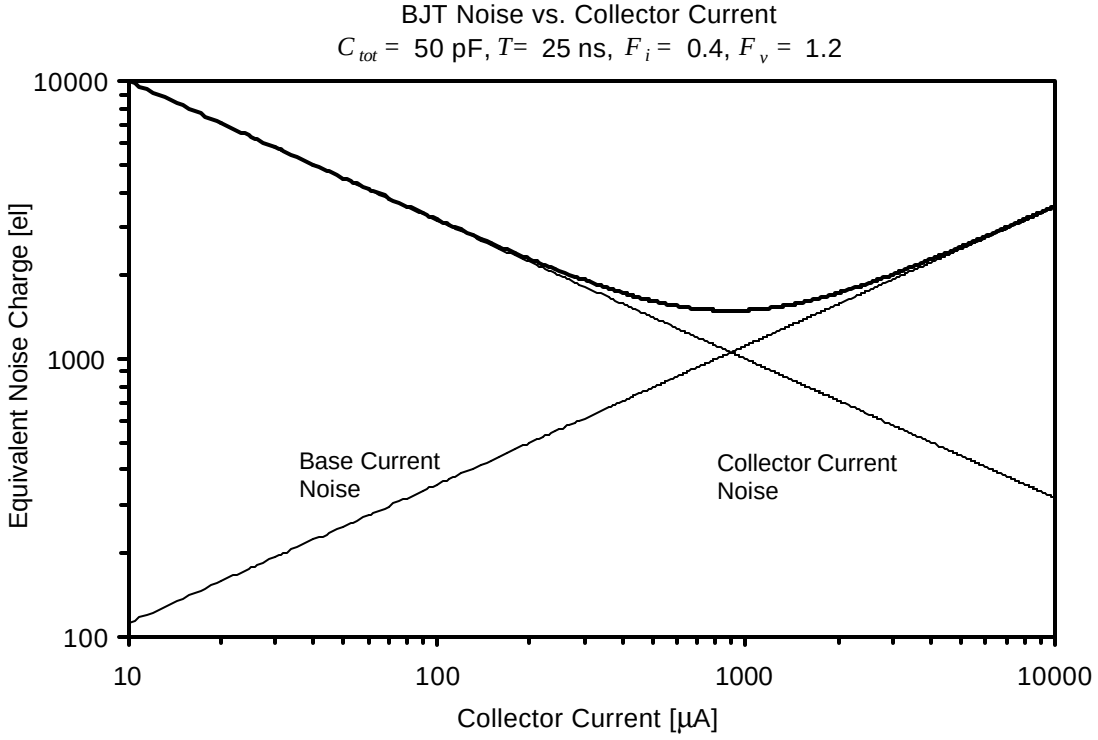


T= 10 ns



Increasing the capacitance at the input shifts the collector current noise curve upwards, so the noise increases and the minimum shifts to a higher current.





Simple Estimate of obtainable BJT noise

For a CR-RC shaper

$$Q_{n,\min} = 772\left[\frac{el}{\sqrt{pF}}\right]\cdot\frac{\sqrt{C}}{\sqrt[4]{b_{DC}}}$$

obtained at $$I_c = 26\left[\frac{mA\cdot ns}{pF}\right]\cdot\frac{C}{t}\sqrt{b_{DC}}$$

Since typically $b_{DC} \approx 100$, these expressions allow a quick and simple estimate of the noise obtainable with a bipolar transistor.

Note that specific shapers can be optimized to minimize either the current or the voltage noise contribution, so both the minimum obtainable noise and the optimum current will be change with respect to the above estimates.

The noise characteristics of bipolar transistors differ from field effect transistors in four important aspects:

1. The equivalent input noise current cannot be neglected, due to base current flow.
2. The total noise does not decrease with increasing device current.
3. The minimum obtainable noise does not depend on the shaping time.
4. The input capacitance is usually negligible.

The last statement requires some explanation.

The input capacitance of a bipolar transistor is dominated by two components,

1. the geometrical junction capacitance, or transition capacitance C_{TE}, and
2. the diffusion capacitance C_{DE}.

The transition capacitance in small devices is typically about 0.5 pF.

The diffusion capacitance depends on the current flow I_E through the base-emitter junction and on the base width W, which sets the diffusion profile.

$$C_{DE} = \frac{\partial q_B}{\partial V_{be}} = \frac{q_e I_E}{k_B T}\left(\frac{W}{2D_B}\right) \equiv \frac{q_e I_E}{k_B T} \cdot \frac{1}{w_{Ti}}$$

where D_B is the diffusion constant in the base and w_{Ti} is a frequency that characterizes carrier transport in the base. w_{Ti} is roughly equal to the frequency where the current gain of the transistor is unity.

Inserting some typical values, I_E=100 µA and w_{Ti} =10 GHz, yields C_{DE}= 0.4 pF. The transistor input capacitance $C_{TE}+C_{DE}$= 0.9 pF, whereas FETs providing similar noise values at comparable currents have input capacitances in the range 5 – 10 pF.

Except for low capacitance detectors, the current dependent part of the BJT input capacitance is negligible, so it will be neglected in the following discussion. For practical purposes the amplifier input capacitance can be considered constant at 1 ... 1.5 pF.

This leads to another important conclusion.

Since the primary noise parameters do not depend on device size and there is no significant linkage between noise parameters and input capacitance

- Capacitive matching does not apply to bipolar transistors.

Indeed, capacitive matching is a misguided concept for bipolar transistors. Consider two transistors with the same DC current gain but different input capacitances. Since the minimum obtainable noise

$$Q_{n,\min}^2 = 4k_B T \frac{C}{\sqrt{b_{DC}}} \sqrt{F_i F_v} \ ,$$

increasing the transistor input capacitance merely increases the total input capacitance C_{tot} and the obtainable noise.

When to use FETs and when to use BJTs?

Since the base current noise increases with shaping time, bipolar transistors are only advantageous at short shaping times.

With current technologies FETs are best at shaping times greater than 50 to 100 ns, but decreasing feature size of MOSFETs will improve their performance.